**Physicalism Requires Functionalism:**

**A New Formulation and Defense of the Via Negativa**

Justin Tiehen

University of Puget Sound

How should ‘the physical’ be defined for the purpose of formulating physicalism? In this paper I defend a version of the via negativa according to which a property is physical just in case it is neither fundamentally mental nor possibly realized by a fundamentally mental property. The guiding idea is that physicalism requires functionalism, and thus that being a type identity theorist requires being a realizer-functionalist. In §1 I motivate my approach partly by arguing against Jessica Wilson’s no fundamental mentality constraint. In §2 I set out my preferred definition of ‘the physical’ and make my case that physicalism requires functionalism. In §3 I defend my proposal by attacking the leading alternative account of ‘the physical,’ the theory-based conception. Finally, in §4 I draw on my definition, together with Jaegwon Kim’s account of intertheoretic reduction, to defend the controversial view that physicalism requires a priori physicalism.

Physicalism is the thesis that nothing exists over and above the physical. In order to assess whether the thesis is true or false, we first need to get clear on exactly what it means. How should the ‘nothing over and above’ clause be understood? Just what is meant by ‘the physical’? These two questions are generally treated separately but I say this is a mistake. In what follows I defend a version of the *via negativa* account of ‘the physical’ that draws in part on a *realizationist* understanding of ‘nothing over and above.’ In slogan-form, the guiding idea is that *physicalism requires functionalism*.

I motivate my approach in §1, partly by arguing against Jessica Wilson’s *no fundamental mentality* constraint, which I show wrongly entails that certain realizer-functionalist views are inconsistent with physicalism. §2 is where I set out my preferred definition of ‘the physical.’ I argue that it is an advantage of my definition that when it is combined with a realizationist understanding of ‘nothing over and above,’ the result is a theoretically unified understanding of physicalism. In §3 I further clarify and defend my proposal by advancing a new objection against the chief rival to the via negativa, the *theory-based conception* of ‘the physical.’ Finally, in §4 I draw on my definition of ‘the physical,’ together with Jaegwon Kim’s account of intertheoretic reduction, to defend the controversial view that physicalism requires *a priori physicalism*.

**1. The Type Identity Theory and Fundamental Mentality**

According to defenders of the via negativa, ‘the physical’ should be characterized negatively rather than positively, it should be defined in terms of what it is not rather than what it is.[[1]](#footnote-1) On the simplest version of the approach, ‘the physical’ is defined as the *nonmental*; physicalism is then the doctrine that nothing exists over and above the distribution of nonmental properties. Unfortunately, as Barbara Montero and Daniel Stoljar have observed, this simple version is fatally flawed: it is unable to accommodate the *psychophysical type identity theory*, the view that every instantiated mental property is identical with some physical property.[[2]](#footnote-2)

If ‘the physical’ simply means the nonmental, then what the type identity theory asserts is equivalent to the claim that mental properties are not mental, a contradiction. Now, even if you reject the type identity theory you should agree that it is not this easy to establish the view’s incoherence. What we have is thus a reductio of the simple version. The question then becomes whether some more sophisticated version of the via negativa can avoid the reductio.

To see a more sophisticated version, consider a proposal due to Jessica Wilson. Wilson suggests that a proper definition of ‘the physical’ should include a no fundamental mentality (NFM) constraint, requiring (inter alia) that fundamental physical properties must not be mental.[[3]](#footnote-3) Wilson is not fully explicit about how she understands the notion of a fundamental property, but I will assume we can think of them at least broadly along the lines that David Lewis understands *perfectly natural properties*: they render their instances perfectly similar in some respect, they are not at all disjunctive or gruesome, they can enter into laws and causal relations, and nothing exists over and above the distribution of such properties (in the same sense of this phrase that figures in the formulation of physicalism).[[4]](#footnote-4) This is not meant as an exhaustive account of what fundamental properties are, but hopefully it gives a sense.

As a response to the problem facing the simple version of the via negativa, I take Wilson’s idea to be that type identity theorists can accept the NFM-constraint without contradiction because those neural properties that are plausible candidates for identification with mental properties are not fundamental. In the stock example, identity theorists propose that pain=firing C-fibers. If we assume that firing C-fibers is not fundamental—suppose that the fundamental properties are instead things like being an electron, being a down quark, etc.—and if we assume in addition that firing C-fibers nevertheless counts as a (non-fundamental) physical property in some broad sense, then the psychophysical type identity theorist who identifies pain with firing C-fibers can accept the NFM-constraint and take it partly to define ‘the physical.’

I believe that Wilson is absolutely right to focus on fundamental mentality. Following the lead of several philosophers—including not just Wilson but also Kim, Montero, Keith Campbell, and David Chalmers—I regard the rejection of fundamental mentality as a core commitment of physicalism.[[5]](#footnote-5) Indeed, my own proposal below (§2) aims to develop this thought. But there is a question of how the notion of fundamental mentality should be understood, and I say that Wilson has understood it in the wrong way for the purpose of defining ‘the physical’ and formulating physicalism.

For Wilson, the crucial issue is whether there are properties that are *both* fundamental *and* mental. When we focus on just those properties that are fundamental, a property’s being mental precludes it from also being physical, according to the NFM-constraint. What physicalists must deny, on Wilson’s view, is that any of the fundamental properties instantiated in the actual world are mental properties. But this is too strong: there are perfectly good physicalist views maintaining that some fundamental mental properties are instantiated, and thus violating the NFM-constraint.

To make my case, consider the mind-body theory known as *realizer-functionalism*.[[6]](#footnote-6) The view is familiar, but because of the pivotal role it plays in several arguments that follow, I review it here at some length. Like functionalists of all sorts, realizer-functionalists hold that mental terms are implicitly defined by their place in some (perhaps folk, perhaps empirical) psychological theory. Call the theory ‘*T*M’ (‘M’ for mental). The characteristic clauses of *T*M describe the causal relations that mental states enter into. Employing the method of Ramseyfication made familiar by Lewis, we represent the postulate of *T*M—a (perhaps infinitely long) sentence entailing all the logical consequences of *T*M—as *T*M(m1,…, mn), where m1,…, mn are all the mental terms used in *T*M, written as names for mental properties.[[7]](#footnote-7) Replacing these mental terms with existentially bound variables yields the Ramsey sentence of *T*M: ∃x1,…,∃xn*T*M(x1,…, xn). An *n*-tuple of properties satisfies the Ramsey sentence of *T*M at a world *w* just in case the members of that *n*-tuple enter into the pattern of relations at *w* that *T*M says mental properties enter into here in the actual world.

Continuing with the Ramsey-Lewis method, we say that for any possible world *w*, if the Ramsey sentence of *T*M is true at *w* then the denotation of the *i*th mental term occurring in *T*M is the property that is the *i*th member of the *n*-tuple satisfying the sentence of *T*M at *w*, while if the Ramsey sentence of *T*M is false at *w* then the mental terms of *T*M are denotationless there.[[8]](#footnote-8) We have thereby specified the extensions of the mental terms of *T*M across all possible worlds, and so if meanings are intensions—understood as functions from worlds to extensions—we have specified the meanings of those terms. And, crucially, we have done so using entirely nonmental language, since all mental terms in *T*M have been replaced by bound variables.

Again, functionalists of all sorts embrace functional definitions broadly along these lines. What distinguishes realizer-functionalism is that it identifies mental properties with the first order properties occupying the functional roles specified by *T*M. This is in contrast with *role-functionalism*, which identifies mental properties with second order “functional” properties of having some first order property or other that occupies such a role. Suppose the pain-role consists in having a property whose instances are typically caused by tissue damage and typically cause winces and groans, and suppose firing C-fibers is the first order property occupying this role. Realizer-functionalists will then identify pain with firing C-fibers, while role-functionalists will identify it with the functional property of having some first order property or other that occupies the role.

Realizer-functionalism is thus a version of the psychophysical type identity theory. Role-functionalism is not. Functional properties are generally taken not to count as physical properties on the basis that it is at least metaphysically possible for them to have nonphysical realizers. For example, it is sometimes said to be metaphysically possible that nonphysical ectoplasm could occupy the pain-role and thus realize the functional property of having some property or other that occupies the role. Still, typical role-functionalists are physicalists who hold that in the actual world, functional properties are always ultimately realized by physical properties, and so are nothing over and above physical properties.

Bringing this to bear on Wilson, I take the most compelling version of realizer-functionalism to be ruthlessly reductionistic, insisting not only that psychology is reducible to neuroscience, but also in the long run absolutely everything is reducible to physics. It operates with an extremely lean ontology, positing no properties other than those described by fundamental physics, and it justifies this leanness on the basis that no other properties are needed to causally explain anything.[[9]](#footnote-9) If there simply are no other properties, then inevitably the occupant of the pain-role will be some property described by fundamental physics, and so realizer-functionalists will identify pain with that property. But then, this form of realizer-functionalism says that some instantiated fundamental property is both mental (since it is identical with pain) and physical (namely, the fundamental physical property occupying the pain-role). This is in violation of Wilson’s NFM-constraint.

Surely, though, the view described is both coherent and consistent with physicalism. There is no incoherence in an ontology that posits no properties beyond those described by fundamental physics. Maybe such a view is ill advised (or maybe not), but we can understand it. Furthermore, surely such an ontology is consistent with physicalism, and indeed represents an especially severe form of physicalism. Continuing on, a proponent of this ontology does not sacrifice her coherence or her physicalist credentials by embracing in addition a realizer-functionalist view of the mind, a view we can plainly understand and that provides a highly deflationary account of the mental. Thus, to defend the NFM-constraint against the present alleged counterexample, Wilson would need to say that physicalism is inconsistent with either (i) realizer-functionalism, (ii) the view that physics describes all properties actually instantiated, or (iii) these two views taken jointly (but not individually). None of these options seems at all promising: realizer-functionalists are otherwise regarded as paradigmatic physicalists, the view that physics can describe all instantiated properties otherwise seems like a paradigmatic (if severe) physicalist position, and there is no plausible reason to think that combining these paradigmatic physicalist views yields an antiphysicalist result.

Given that the view described is coherent and consistent with physicalism, Wilson’s NFM-constraint must be rejected. Even when we restrict our attention to fundamental properties, a property’s being mental does not preclude its also being physical. Physicalists in good standing can hold that fundamental mental properties are instantiated in the actual world, provided they are realizer-functionalists who identify these fundamental mental properties with physical properties.

It may be further clarifying to sketch two versions of such a realizer-functionalist view: one that is more empirically plausible but relies on a controversial metaphysical principle (I prefer this version), and one that does without the principle in question. Starting with the former, consider a realizer-functionalist who identifies pain with firing C-fibers. This psychoneural type identification is all well and good, but presumably firing C-fibers cannot now be treated as an irreducible neural property—after all, what effects does it cause that cannot be explained by fundamental physics? So, the realizer-functionalist must suppose that firing C-fibers can in turn be identified with some underlying physical property, presumably some complex structural property that is “built from” fundamental physical properties.[[10]](#footnote-10) Something like the structural property of having an electron here, a down quark there, and on and on.[[11]](#footnote-11)

But now a problem arises. Why should we believe in the existence of such structural properties, as opposed to believing just in their parts, the fundamental properties from which they are built? After all, structural physical properties do not seem to do any extra causal work that is not done just by their parts acting in concert.[[12]](#footnote-12) And if structural physical properties do not exist, or even if they exist but are epiphenomenal (because causally excluded by their parts), then no such property can occupy the pain-role. In that case, realizer-functionalists will not identify pain with any such structural physical property.

Now, some philosophers respond to this sort of causal exclusion concern by positing *causal overdetermination*. Wholes and parts causally overdetermine their various effects, these philosophers say.[[13]](#footnote-13) Whatever the merits of this response, however, it seems to be unavailable to typical realizer-functionalists, given their own reliance on causal exclusion reasoning to justify the identification of mental with physical properties. If they were to embrace overdeterminationism here, they would lose their own best argument for the type identity theory.[[14]](#footnote-14)

So here is what I recommend instead. Realizer-functionalists should say that the relation between a structural property and the fundamental properties it is built from is that of (perhaps non-mereological) *composition*, and then they should adopt the metaphysical principle that *composition is identity*.[[15]](#footnote-15) If composition is identity, then a structural property cannot be causally excluded by the fundamental physical properties that compose it, for they are one and the same, the whole and its parts. The structural property just is those fundamental properties, taken jointly.[[16]](#footnote-16)

By holding that composition is identity, realizer-functionalists can give a unified response to causal exclusion worries. The save the causal efficacy of mental properties by identifying them with physical properties, and then they save the causal efficacy of those same (structural) physical properties by identifying them with the (causally efficacious) fundamental physical properties from which they are built. In each case, the identity relation is the antidote to the causal exclusion poison.

Continuing on, realizer-functionalists who take this line should then add that structural physical properties *are themselves fundamental*. Write up a list of all the fundamental properties. Once you include on your list all the fundamental properties that compose a given structural property, you thereby include the structural property itself—for they are identical, the whole and its parts. [[17]](#footnote-17) (Compare: once you include H2O on your list of chemical properties, you thereby include water.) Notice that this is no expansion to the lean ontology that realizer-functionalists promised. The structural physical properties we are “adding” to the list of fundamental properties were there all along, since their parts were there. And notice it’s a good thing that structural properties end up qualifying as fundamental, for the key premise in the causal exclusion problem we were confronting is that no properties other than fundamental physical ones are needed for causal explanation, and nowhere have we rejected this premise.[[18]](#footnote-18)

I regard this as the best version of realizer-functionalism, the most consistent version. We might ultimately favor some other mind-body theory, but if the position is so much as coherent, Wilson’s NFM constraint must be rejected. For the view described is a physicalist one, and yet it says that some of the fundamental properties instantiated in the actual world (i.e., certain structural physical properties) are mental.[[19]](#footnote-19)

Suppose you don’t buy this though, perhaps because you doubt the coherence of composition as identity.[[20]](#footnote-20) I have a backup argument. Assuming realizer-functionalism, we can imagine further empirical investigation showing that the property occupying the pain-role is not firing C-fibers, or any other neural property, or any biological property, but instead is some (non-structural) property described by fundamental physics—a certain charge property, say. This seems empirically unlikely, but it is conceivable (it cannot be ruled out a priori).[[21]](#footnote-21) Surely, this empirical result would be consistent with physicalism. After all, what we are imagining is not that the charge property in question has an intrinsic mental essence, a qualitative character that physics fails to describe. That would take us in the direction of panpsychism, a view that very plausibly *is* inconsistent with physicalism (see §3). Rather, what we are imagining is that the realizer-functionalist’s deflationary account of the mental is correct, and so there is nothing more to pain than instantiating a property that occupies the right causal role; and then we are imagining in addition that some property described by fundamental physics has the right causal powers to occupy this role. Nothing is hidden to physics. Why on earth would a physicalist need to deny that a property described by fundamental physics has these particular powers?

The argument can be reframed in metaphysical rather than epistemological terms. Consider a possible world *w* in which the given charge property occupies the pain-role. If the laws of nature are metaphysically contingent—as in fact the leading realizer-functionalists hold[[22]](#footnote-22)—there is bound to be such a world; indeed, for any property (or at least any property that can enter into causal relations at all) and for any causal role, there is some world where that property occupies that role. According to realizer-functionalism, at *w* the charge property in question is identical with pain (i.e., is the denotation of ‘pain’),[[23]](#footnote-23) and so some property that is both fundamental and mental is instantiated. And yet, the world as described is consistent with physicalism. Therefore, we must reject Wilson’s NFM-constraint.

**2. Physicalism Requires Functionalism**

Physicalism, I have argued, is consistent with the instantiation of fundamentalmental properties—that is, properties that are both mental and fundamental. What it is not consistent with, I say instead, is the instantiation of *fundamentally* mental properties—that is, properties whose status *as mental* is primitive, unanalyzable. To develop this thought further, so that we can use it in a definition of ‘the physical,’ we need an analysis of what it means to be fundamentally mental. Considering that realizer-functionalism is the view that has been posing a problem, I propose as a bold (if somewhat flatfooted) hypothesis that we can understand the notion entirely in functionalist terms.

When realizer-functionalists propose the identification of pain with firing C-fibers, they provide an account of that by virtue of which firing C-fibers qualifies as pain, and so as mental; an account of how various (non-mental) causal relations that the property enters into ground its status as pain. Or, to engage in semantic ascent, they provide an account of what it is about firing C-fibers that makes it the denotation of ‘pain.’ This account comes in the form of their functional definition of ‘pain.’

Let’s say that a property is **fundamentally mental** just in case it is (i) mental and (ii) such that there is no analysis in the form of a functional definition of that by virtue of which it qualifies as the mental property it is (is the denotation of the mental term it is). Potential examples include qualia as antiphysicalists like Chalmers conceive them.[[24]](#footnote-24) Chalmersian qualia are supposedly mental, fundamental (they are something “over and above” physical properties), and defy functional definition—they do not qualify as the mental properties they are by virtue of occupying some functional role, but instead are primitively mental, fundamentally mental. These are the sorts of properties that physicalists must deny are instantiated in the actual world.

With this notion of fundamentally mental properties in place, I propose the following via negativa approach to defining ‘the physical’: **a property is physical just in case it is neither fundamentally mental nor possibly realized by a fundamentally mental property**.[[25]](#footnote-25) Physicalism is then the thesis that nothing exists over and above the distribution of properties that are physical in this sense.

Reconsider the scenario that posed a problem for Wilson. It involved a property that was both mental and fundamental, but that qualified as mental purely by virtue of satisfying a certain functional definition. My definition entails that such a property counts as physical—it is neither fundamentally mental nor possibly realized by a fundamentally mental property—and thus that the scenario is consistent with physicalism. The correct result.

The clause about possible realization is included in my definition to honor the point made in §1, that functional properties are not physical properties. Antiphysicalists like Chalmers sometimes take qualia to be epiphenomenal here in the actual world, but I assume this is not a metaphysically necessary truth.[[26]](#footnote-26) There are worlds in which fundamentally mental properties cause things, and so occupy various functional roles, and so realize functional properties. Because functional properties have these metaphysically possible nonphysical realizers, my definition entails they do not count as physical. Again, the correct result.

One way to generate support for my proposed definition is by considering that other big question of physicalism: how ‘nothing over and above’ should be understood. It was once common to analyze the notion in terms of supervenience,[[27]](#footnote-27) but this approach has fallen out of favor in recent years. What physicalism plausibly requires is not just that all properties supervene on physical properties, but in addition that the supervenience relations themselves be physicalistically explainable rather than brute.

The philosophers advancing this critique of supervenience-based formulations typically appeal instead to the realization relation, familiar from role-functionalism, to make sense of ‘nothing over and above.’[[28]](#footnote-28) If we understand realization as the relation that obtains between the instance of some first order property occupying a certain functional role and the instance of the second order property of having some first order property or other occupying that role, then realization *entails that* and *explains why* the given second order property supervenes on the first order one.

An implication of embracing this realizationist understanding of ‘nothing over and above’ is that—if we very briefly set aside the type identity theory—being a physicalist *requires* being a kind of role-functionalist, since the notion of realization at the heart of the proposal is just that taken from role-functionalism.Having obtained this result, I now add that an implication of my proposed definition of ‘the physical’ is that being a psychophysical type identity theorist *requires* being a realizer-functionalist, since my definition entails that any mental property that is also physical must not be fundamentally mental, in which case there must be an account in the form of a functional definition of that in virtue of which it qualifies as the mental property it is. Put these two results together and you get the conclusion that *physicalism requires functionalism*, either role-functionalism or realizer-functionalism about the mental.

My proposed definition of ‘the physical’ thus acts in conjunction with the realizationist understanding of ‘nothing over and above’ to impose a *uniform requirement* on physicalists: they must be functionalists. There is room for disagreement as to what sort of functionalist to be, but this is a dispute between parties that agree about much.[[29]](#footnote-29)

*Physicalism requires functionalism* is an incautious statement of the view. Qualification is needed. A thoroughgoing eliminative materialism that denies the existence of all mental states is a physicalist view but not a functionalist one. In response, I restrict the scope of my claim: being a physicalist requires being a functionalist about whatever mental states *you take to exist*. If you don’t take any mental states to exist, you can be a physicalist without being a functionalist of any sort. This doesn’t seem ad hoc; it seems like a natural restriction to make. Next, consider behaviorism, a form of physicalism often contrasted with functionalism. Here I think the thing to say is that the operative notion of functional definition is broad enough to include behaviorist definitions of mental terms as a special case: they are functional definitions that define mental terms one-by-one rather than all at once, and so that don’t specify mental states in terms of how they causally interact with other mental states.[[30]](#footnote-30)

If there are yet other physicalist mind-body options worth considering, I claim in advance that they can be handled in one of these two ways: either by imposing a non-ad hoc restriction on the claim that physicalism requires functionalism, or by fitting the overlooked alternative into the framework of functional definitions, and thus by treating it as a special case of functionalism.[[31]](#footnote-31) The claim then is that physicalism requires functionalism, given these qualifications.

It is a consideration in favor of my proposed definition of ‘the physical’ that it achieves this kind of theoretical unity in our understanding of physicalism. The proposal has further virtues as well, however. A definition of ‘the physical’ should be judged on whether it sheds light on various foundational issues concerning physicalism. Does it clarify the terms of some dispute, or obscure them? In the remaining sections of the paper I further defend my proposal by showing what it clarifies. In the process, I respond to potential objections.[[32]](#footnote-32)

**3. The Theory-Based Conception**

I take the single leading alternative to the via negativa to be the view Stoljar calls the *theory-based conception*.[[33]](#footnote-33) It says ‘the physical’ should be defined in terms of the comprehensive theory of the science of physics. Let’s say that a property designated by a term from this theory is a *physicalT* *property* (‘T’ for theory), and that *physicalismT* is the thesis that nothing exists over and above the distribution of physicalT properties. Defenders of the theory-based conception say that physicalism should be equated with physicalismT. I deny this; I reject the theory-based conception.[[34]](#footnote-34)

Some critics of the theory-based conception make their case by appealing to *Hempel’s dilemma*.[[35]](#footnote-35) The dilemma says that if ‘the physical’ is defined in terms of the entities described by contemporary physics, the resulting physicalistT doctrine is very likely false—contemporary physics is very likely incomplete—while if it is defined in terms of the entities described by some future, idealized physics, physicalismT is empty—who knows what future physics will contain? The purported upshot is that there is no promising way to develop the theory-based conception.

My objection is independent of Hempel’s dilemma however. Grant defenders of the theory-based conception that they have some adequate reply to the dilemma without worrying too much about how the details go. Even so, the conception fails to capture what really maters in the debate regarding physicalism, the exact relation between mind and body. The truth of physicalismT is neither necessary nor sufficient for the truth of physicalism (properly understood).

First, it is not *necessary*. Consider a scenario in which chemical properties are emergent, they are something over and above the properties described by physics, but such emergence occurs nowhere else in nature. In particular, minds are not emergent. Realizer-functionalism is the correct mind-body theory, and mental states are identical with neural states in just the way realizer-functionalists suppose. Then physicalismT is false and yet, I claim, physicalism (properly understood) is true. After all, the scenario gives typical self-described physicalists what they want most; it vindicates their view of the mind-body relation while deviating on a topic they are not especially invested in, the exact relation between physics and chemistry. (There is a reason that discussions of physicalism focus so much on minds, so little on chemistry.) My definition of ‘the physical’ delivers the correct result: the scenario involves no fundamentally mental properties, and so it is consistent with physicalism as I define it.[[36]](#footnote-36)

Next, it is not *sufficient*. To set up my argument, consider the question of how the theoretical terms occurring in the (perhaps contemporary, perhaps future) comprehensive theory of physics, *T*P (‘P’ for physics), should be defined. I suggest a realizer-functionalist treatment: Ramsify *T*P, replacing its theoretical terms with existentially bound variables, and then use the resulting Ramsey sentence to assign extensions to those terms across all possible worlds.[[37]](#footnote-37) For example, if ‘being an electron’ is defined as the property whose instances are attracted to protons, repel other electrons, annihilate positrons upon collision, and so on, then we take ‘being an electron’ to denote whatever property occupies this role.[[38]](#footnote-38)

With this account in place, consider three different possible worlds. At each world, the Ramsey sentence of *T*P is true and every instantiated fundamental property is denoted by some theoretical term of physics. But the underlying natures of the denoted properties are different across the worlds. At the first world, none of the instantiated fundamental properties is fundamentally mental. Perhaps no mental properties at all are instantiated at the world, or perhaps they are instantiated but not fundamental, or perhaps they are instantiated and fundamental but not fundamentally mental (as on the realizer-functionalist view described in §1). In that case, physicalism is true at the world regardless of whether ‘the physical’ is defined using my proposal from §2 or using the theory-based conception. Accordingly, we shall call this *Physicalism World.* Perhaps this is how the actual world is, and so perhaps physicalismT is in fact true. I am willing to suppose that it is for the sake of my present objection.

At the second world, every instantiated fundamental property is fundamentally mental. Again, you can think of these properties as qualia as antiphysicalists like Chalmers conceive them. At this world, qualia enter into various causal and other relations with one another, fitting into just the pattern needed for the Ramsey sentence of *T*P to be true, ensuring that the theoretical terms of physics have denotations at the world. So for example, some fundamentally mental property occupies the electron-role and so is the denotation of ‘being an electron’ at the world.[[39]](#footnote-39) Call this place *Idealism World*. If physicalism is the thesis that nothing exists over and above the physical, idealism presumably is the thesis that nothing exists over and above the mental. This thesis is true at Idealism World.

Idealism World represents a kind of panpsychist scenario in which every fundamental physicalT property is fundamentally mental. Lewis at one point entertains the thought that this is how the actual world is, while Chalmers has advanced arguments in favor of such a view without fully committing himself to it.[[40]](#footnote-40) It offers one way to understand Russell’s position in *The Analysis of Matter* that physics tells us about the relational nature of things but not about their intrinsic makeup, and that this intrinsic makeup consists in something like qualia.[[41]](#footnote-41)

At the third possible world, two completely different families of fundamental properties are instantiated. One family consists of properties that are fundamentally mental—some of the same Chalmersian qualia found at Idealism World. The other family consists of properties that are not mental at all—some of the same fundamental properties found at Physicalism World. Taken together, these two families of fundamental properties enter into the pattern of relations described by *T*P, and so every property belonging to either family is denoted by some theoretical term of physics. Perhaps some fundamentally mental property occupies the electron-role, while some property that is not mental at all occupies the positron-role. Call this *Dualism World*, since the two families of fundamental properties are so different.

If we use my definition of ‘the physical,’ physicalism is false at both Idealism World and Dualism World since each contains the instantiation of fundamentally mental properties. In contrast, physicalismT is true at each world, since every fundamental property instantiated at either world is physicalT (is the denotation of some theoretical term of *T*P). The question then is what we want out of a definition of ‘the physical.’ Do we want a definition on which physicalism comes out true at Idealism World and Dualism World, or a definition on which it comes out false?

There are obvious and compelling reasons to prefer the latter. A definition should clarify rather than obscure, it should capture what we care about most in a debate. But if we adopt the theory-based conception we will end up classifying Chalmers as a physicalist (physicalistT), or at least as holding a view consistent with physicalism. We will then need to come up with some new term for philosophers like Lewis who reject Chalmers’ view of consciousness and hold some more deflationary account. This is silly. We don’t need a new word: just keep calling Lewis a physicalist, Chalmers an antiphysicalist. But in that case, we must reject the theory-based conception.

It gets worse for the theory-based conception. As Idealism World demonstrates, idealism and physicalismT are consistent. But surely it is more confusing than helpful to have a definition of ‘the physical’ on which even after physicalism is established, it remains an open question whether idealism is true. Yes, I know you are a physicalist—but are you *also* an idealist? This should not be an open question. But in that case, we must reject the theory-based conception.[[42]](#footnote-42)

Some philosophers have argued that physicalism is consistent with panpsychism: physicalism is consistent with *some* entities having minds, after all, and so why wouldn’t it be consistent with *all* entities having minds?[[43]](#footnote-43) I endorse Wilson’s response to this: what physicalism is incompatible with is fundamental mentality, and while the view that some entities have minds does not require fundamental mentality, the leading versions of panpsychism do. Again, though, my understanding of fundamental mentality is different than Wilson’s.[[44]](#footnote-44)

Continuing on, physicalismT is consistent with a certain kind of dualism, as Dualism World demonstrates. This is another embarrassment for the theory-based conception. Once you posit the instantiation of fundamentally mental properties, you lose your physicalist credentials. You do not gain them back by positing in addition the instantiation of other properties, properties that are not physicalistically objectionable. But in that case, the theory-based conception mishandles Dualism World.

The core problem we are running up against is that the theory-based conception fails to capture whether the world is fundamentally physical and only derivatively mental, or fundamentally mental and only derivatively physical, or fundamentally a bit of both. That is, it fails to capture what the mind-body relation really is. This is the heart of the traditional dispute between physicalists, idealists, and dualists, but physicalismT is entirely neutral on the question. Physicalism should not be neutral; it should be a doctrine that takes a stand. But in that case, we must reject the theory-based conception.

**4. A Posteriori Type Physicalism and the Asymmetry Constraint**

Physicalism requires functionalism, and so being a psychophysical type identity theorist requires being a realizer-functionalist. Or so I have argued. But several contemporary identity theorists disagree, embracing type identities while eschewing functional definitions.[[45]](#footnote-45) One way to motivate their view is by reflecting on certain epistemic arguments against physicalism; here, we will focus on Chalmers’s conceivability argument.

Let *P* be the conjunction of all physical truths, and *M* the mental truth that Chalmers is conscious. Then, consider the following argument:

(P1): *P*&*~M* is conceivable.

(P2): If *P*&*~M* is conceivable, *P*&~*M* is logically possible.

(P3): If *P*&~*M* is logically possible, physicalism is false.

(C): Physicalism is false.[[46]](#footnote-46)

*P*&*~M* is the hypothesis that Chalmers is a zombie: all the physical truths are just as they actually are but Chalmers is not phenomenally conscious. (P1) states that this zombie hypothesis is conceivable, where a proposition is conceivable just in case its negation is not knowable a priori.

Realizer-functionalists must deny (P1). If all mental terms are functionally defined, it is knowable a priori that if the physical truths of the actual world obtain then Chalmers is conscious, since he instantiates various physical properties occupying the right functional roles for consciousness. Realizer-functionalism is thus a form of *a priori type physicalism*, that is, a type physicalist view that asserts that the physical truths *a priori entail* all truths, including that Chalmers is conscious.[[47]](#footnote-47) In the judgment of many philosophers—including both physicalists and antiphysicalists—(P1) has a great deal of intuitive plausibility, however. If realizer-functionalism requires the rejection of (P1), some will regard this as sufficient reason to reject realizer-functionalism.[[48]](#footnote-48)

In contrast, *a posteriori type physicalists*, who accept psychophysical property identities but deny that mental terms are functionally definable, grant (P1) but deny (P2), maintaining that conceivability does not entail possibility.[[49]](#footnote-49) The zombie hypothesis cannot be ruled out a priori, they admit, but this epistemic premise does not support the metaphysical conclusion that physicalism is false. A posteriori type physicalists characteristically go on to offer alternative accounts of the semantics of mental terms,[[50]](#footnote-50) but here we will not worry about the details of how such alternative accounts go other than to suppose that they do not take the form of functional definitions and do not otherwise entail the negation of (P1).

In response to a posteriori type physicalism, I stick to my guns: physicalism requires functionalism, I reassert. Consequently, I deny that a posteriori type physicalism is a coherent physicalist view; to be a type physicalist, you must be an a priori type physicalist, I say.[[51]](#footnote-51) To set up my argument for this conclusion, it will be helpful to consider Kim’s influential account of *intertheoretic reduction*.[[52]](#footnote-52)

Roughly stated, Kim defends a realizer-functionalist model consisting of three steps. (1) Functionally define the terms of the theory to be reduced—the theory may be psychology, say, or classical thermodynamics. (2) Through empirical investigation, discover which properties described by the reducing theory occupy the functional roles specified by the reduced theory—the reducing theory may be neuroscience, say, or statistical mechanics. (3) Identify each property described by the reduced theory with the property described by the reducing theory that occupies the appropriate functional role—identify pain with firing C-fibers, say, or heat with molecular motion.[[53]](#footnote-53)

I endorse Kim’s account, although I think I understand it slightly differently than he does. Kim describes step (1), in which the terms of the reduced theory are functionally defined, as merely “priming” the reduction, preparing the way for it.[[54]](#footnote-54) The suggestion seems to be that the reduction itself takes place entirely in the final step (3), the property identifications. In that case, if the property identifications could somehow be secured in some other way, the reduction would still go through.

I deny this. I say the functional definitions of step (1) do not merely clear the ground for reduction, but rather are partly *constitutive* of it. The property identities of step (3) are perhaps necessary for reduction, but they are insufficient. Last section’s discussion of Idealism World and Dualism World illustrate why. Suppose I grant that every mental property is identical with some physicalT property. Does it follow that psychology is reducible to physics? No. If the properties in question are fundamentally mental, and physicalT by virtue of occupying the right functional roles, then psychology is irreducible to physics. And Kim’s model agrees—at least if we take its step (1) seriously, as a necessary condition for reduction. If anything, we should say that at Idealism World physics is reducible to psychology, while at Dualism World neither science is reducible to the other (physics is not reducible to psychology there, given that some physicalT properties are not mental).

There is a familiar puzzle for accounts of intertheoretic reduction that assign a central role to property identifications: identity is a symmetric relation, while reduction is asymmetric. If classical thermodynamics is reducible to statistical mechanics, then statistical mechanics is not likewise reducible to thermodynamics. Something thus must be added to property identities to capture this asymmetry. One thought that might occur to you is to add that every property in the reduced theory is identical with some property in the reducing theory, but not vice versa. But this will not do. At Dualism World every fundamentally mental property is identical with some physicalT property, but not vice versa, and yet psychology is not reducible to physics there. What must be added instead, I claim, is that the terms of reduced theory must be functionally defined using vocabulary from the reducing theory (together with topic-neutral vocabulary).[[55]](#footnote-55)

With Kim’s model of reduction in mind, reconsider a posteriori type physicalism. The view’s critics typically argue that in the absence of functional definitions of mental terms, there is no way to obtain warrant for the proposed psychophysical type identities, as opposed to the dualist counterhypothesis that mental and physical properties are distinct but nomically correlated.[[56]](#footnote-56) The view’s defenders then typically respond that such identities can be warranted in some other way, perhaps abductively.[[57]](#footnote-57) I want to bypass this debate, which has reached a stalemate. Grant a posteriori type physicalists the property identities they want. My claim is that such identities, in the absence of functional definitions, do not provide the asymmetry needed for psychology to be reducible to neuroscience, and in connection do not provide the asymmetry between the mental and the physical that physicalism demands.

To illustrate the idea, consider a scenario in which physicalism is false. Suppose pain is a fundamentally mental property, a Chalmersian quale. It follows that ‘pain’ is not functionally definable. Next, suppose that ‘firing C-fibers,’ as a theoretical term of neuroscience, *is* functionally definable, and that the term denotes whatever property occupies the firing C-fibers-role. Finally, suppose that empirical investigation shows that pain, the fundamentally mental property, occupies this role. Then pain=firing C-fibers even though physicalism (by assumption) is false. The scenario is like those represented by Idealism World and Dualism World, except that it involves the functional definition of neuroscientific terms rather than terms from physics.

The challenge for a posteriori type physicalists is to say how this antiphysicalist scenario differs from the actual world as they conceive it. We have supposed that ‘pain’ is not functionally definable, just as they insist. We have supposed that ‘firing C-fibers’ is functionally definable, which they could try to deny, but this seems unpromising. After all, if the functional definition of ‘firing C-fibers’ were inadequate in just the way the functional definition of ‘pain’ is supposed to be, it would become mysterious why the conceivability argument regarding consciousness is so much more compelling than an analogous conceivability argument regarding neural states.[[58]](#footnote-58) Finally, we have supposed that pain occupies the firing C-fibers-role, but a posteriori type physicalists surely will accept this since they hold that pain=firing C-fibers and that firing C-fibers occupies the firing C-fibers-role. We have given a posteriori type physicalists so much of what they want, and yet the scenario described is physicalistically unacceptable.

I suspect most a posteriori type physicalists think their physicalist credentials are established simply by the psychophysical property identities they posit. But this is not so: again, identity is symmetric, while physicalism requires an asymmetry between the mental and the physical. And again, it will not do to add that every mental property is identical with some neural property (or, alternatively, some property described by physics) but not vice versa. This fails to provide the right kind of asymmetry, as Dualism World demonstrates.

Where are they going to get the needed asymmetry without appealing to functional definability? I don’t know; I see no promising options. I cannot prove there is no way to get it, but I can further clarify my skepticism by focusing on one option that will bring our discussion full circle. A posteriori type physicalists could hold that neural properties like firing C-fibers are structural properties that are composed of lower level properties (ultimately, properties described by physics), and that these lower level properties are not mental at all. In that case, it is not the symmetric identity relation between pain and firing C-fibers that establishes the physicalist credentials of the view, but the asymmetric composition relation that holds between pain (a.k.a., firing C-fibers) and the non-mental properties it is built from.[[59]](#footnote-59)

But this response quickly runs into trouble: just what is this composition relation that is being invoked? If composition is identity (§1), the response fails to deliver the asymmetry sought, since identity is symmetric. In that case the component properties are (collectively) mental after all, since they are (collectively) identical with the given structural property, and the structural property is identical with pain. Moreover, in that case pain will plausibly qualify as a fundamental property (for the reasons set out in §1), and so a posteriori type physicalists will need to explain how their brand of physicalism can allow for the instantiation of fundamental mental properties. And in explaining this, they cannot help themselves to my own account of how this would work (§1), since my account made use of the realizer-functionalist framework they reject.[[60]](#footnote-60)

So suppose composition is not identity. Some philosophers have argued that composition should be understood as involving a kind of emergence, of the sort that antiphysicalist philosophers like the British Emergentists embraced.[[61]](#footnote-61) A whole, they say, is not identical with its parts; it is a distinct entity, something over and above the parts, something emergent. Obviously a posteriori type physicalists cannot accept this. It would mean that pain is something over and above its supposed nonmental components, an emergent, fundamentally mental property that qualifies as firing C-fibers by virtue of occupying the right functional role. This is an emergentist version of the antiphysicalist scenario we were imagining, the one we had trouble distinguishing from the a posteriori type physicalist view.

What is needed is a view of composition according to which wholes are not identical with their parts, but they also are nothing over and above their parts. Furthermore, the operative notion of ‘nothing over and above’ cannot be cashed out in terms of supervenience, for there are compelling objections to such supervenience-based formulations (§2). And it cannot be cashed out in terms of realization, at least not if realization is understood in terms of the framework of functionalism and functional definitions (§2), for a posteriori type physicalists deny that ‘pain’ is functionally definable.

Summarizing: to establish their physicalist credentials, a posteriori type physicalists who take the line we are envisioning need a new account of ‘nothing over and above,’ one that is different from any of the leading accounts of ‘nothing over and above’ found in the physicalism literature, and one that avoids the twin perils of identity and emergence. This is a tall order. I doubt it can be done.[[62]](#footnote-62)

In light of this, I deny that it is an embarrassment for my proposed definition of ‘the physical’ that it makes no room for a posteriori type physicalism. A definition of ‘the physical’ should shed light on foundational debates concerning physicalism. In the present case, my definition focuses attention on the asymmetry constraint that a posteriori type physicalists must somehow satisfy. Chalk this up as a further virtue of the definition, I say.[[63]](#footnote-63)

WORKS CITED

Armstrong, D. M. 1968. *A Materialist Theory of Mind*. New York: Humanities Press.

Armstrong, D. M. 1978. *A Theory of Universals: Volume 2: Universals and Scientific Realism*.

Cambridge: Cambridge University Press.

Armstrong, D. M. 1983. *What is a Law of Nature*? Cambridge: Cambridge University

Press.

Armstrong, D. M. 1986. “In Defense of Structural Universals,” *Australasian Journal of*

*Philosophy*, 64.1: 85-88.

Baxter, D. 1988a. “Identity in the Loose and Popular Sense,” *Mind*, 97: 575-582.

Baxter, D. 1988b. “Many-One Identity,” *Philosophical Papers*, 17.3: 193-216.

Block, N. and R. Stalnaker. 1999. “Conceptual Analysis, Dualism, and the Explanatory

Gap,” *Philosophical Review*, 108: 1-46.

Campbell, K. 1997. “Review of *Physicalism: The Philosophical Foundations*,” *Philosophy and*

*Phenomenological Research*, 57: 223-226.

Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford:

Oxford University Press.

Chalmers, D. J. 2003. “Consciousness and Its Place in Nature,” in S. Stich and T. Warfield

(eds.) *Blackwell Guide to the Philosophy of Mind*. Blackwell.

Chalmers, D. J. 2012. *Constructing the World*. Oxford: Oxford University Press.

Chalmers, D. J. Forthcoming. “Panpsychism and Panprotopsychism,” URL =

<consc.net/papers/panpsychism.pdf>.

Chalmers, D. J. and F. Jackson. 2001. “Conceptual Analysis and Reductive Explanation.”

*Philosophical Review*, 110: 315-361.

Chomsky, N. 1993. *Language and Thought*. Rhode Island: Moyer Bell.

Chomsky, N. 1995. “Language and Nature,” *Mind*, 104: 1-61.

Davidson, D. 1970. “Mental Events,” in L. Foster and J. W. Swanson (eds.), *Experience*

*and Theory*. Amherst: Massachusetts University Press.

Dowell, J. 2006. “The Physical: Empirical, Not Metaphysical,” *Philosophical Studies*, 131.1:

25-60.

Fodor, J. 1974. “Special Sciences,” *Synthese*,28: 77-115.

Hawthorne, J. 2001. “Causal Structuralism,” *Philosophical Perspectives*, 15: 361-378.

Hempel, C. G. 1969. “Reduction: Ontological and Linguistic Facets,” in P. Suppes, S.

Morgenbesser, and M. White (eds.), *Philosophy, Science, and Method: Essays in*

*Honor of Ernest Nagel*. St. Martin’s.

Hempel, C. G. 1980. “Comments on Goodman’s *Ways of Worldmaking*,” *Synthese*, XLV:

193-199.

Hill, C. 1991. *Sensations: A Defense of Type Materialism*. Cambridge University Press.

Hill, C. 1997. “Imaginability, Conceivability, Possibility, and the Mind-Body Problem,”

*Philosophical Studies*, 87: 61-85.

Horgan, T. 1993. “From Supervenience to Superdupervenience,” *Mind*, 102: 555-586.

Horgan, T. and J. Tienson. 2001. “Deconstructing New Wave Materialism,” in C. Gillett

and B. M. Loewer (eds.), *Physicalism and Its Discontents*. Cambridge University Press.

Horgan, T. 2010. “Materialism, Minimal Emergentism, and the Hard Problem of

Consciousness,” in R. C. Koons and G. Bealer (eds.), *The Waning of Materialism*. Oxford: Oxford University Press.

Jackson, F. 1982. “Epiphenomenal Qualia,” *Philosophical Quarterly*, 32: 127-136.

Jackson, F. 1996. “Mental Causation: The State of the Art,” *Mind*, 105: 377-413.

Jackson, F. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford:

Oxford University Press.

Jackson, F. 2002. “From Reduction to Type-Type Identity,” *Philosophy and*

*Phenomenological Research*, LXV, 3: 644-647.

Jackson, F. 2003a. “From H2O to Water: The Relevance to A Priori Passage,” in H.

Lillehammer and G. Rodriguez-Pereyra (eds.) *Real Metaphysics*. Routledge.

Jackson, F. 2003b. “Mind and Illusion,” *Royal Institute of Philosophy Supplement*, 53: 251-

271.

Jackson, F. 2007. “A Priori Physicalism,” in B. P. McLaughlin and J. Cohen (eds.)

*Contemporary Debates in Philosophy of Mind*. Blackwell.

Jackson, F., R. Pargetter, and E.W. Prior. 1982. “Functionalism and Type-Type Identity

Theories,” *Philosophical Studies*, 42: 209-225.

Jenkins, C. S. “Is Metaphysical Dependence Irreflexive?” *The Monist*, 94: 267-276.

Kane, R. 1998. *The Significance of Free Will*. Oxford: Oxford University Press.

Kim, J. 1996. *Philosophy of Mind, 1st Edition*. Westview Press.

Kim, J. 1998. *Mind in a Physical World*. Cambridge: MIT Press.

Kim, J. 2002. “Response to Frank Jackson,” *Philosophy and Phenomenological Research*,

LXV, 3: 671-674.

Kim, J. 2005. *Physicalism, Or Something Near Enough*. Princeton University Press.

Kripke, S. A. 1980. *Naming and Necessity*. Oxford: Blackwell.

Lewis, D. K. 1966. “An Argument for the Identity Theory,” *Journal of Philosophy*, 63: 17-

25.

Lewis, D. K. 1969. “Review of *Art, Mind, and Religion*,” *Journal of Philosophy*, 66: 22-27.

Lewis, D. K. 1972. “Psychophysical and Theoretical Identifications,” *Australasian Journal*

*of Philosophy*, 50: 249-258.

Lewis, D. K. 1973. *Counterfactuals*. Cambridge: Harvard University Press.

Lewis, D. K. 1983. “New Work for a Theory of Universals,” *Australasian Journal of*

*Philosophy*, 61: 343-377.

Lewis, D. K. 1986. “Against Structural Universals,” *Australasian Journal of Philosophy*, 64:

25-46.

Lewis, D. K. 1994. “Reduction of Mind,” in S. Guttenplan (ed.), *A Companion to*

*Philosophy of Mind*. Blackwell.

Lewis, D. K. 2009. “Ramseyan Humility,” in D. Braddon-Mitchell and R. Nola (eds.),

*Conceptual Analysis and Philosophical Naturalism*. Cambridge: MIT Press.

Levine, J. 2001. *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.

Loar, B. 1997. “Phenomenal States,” in N. Block, O. Flanagan, and G. Güzeldere (eds.),

*The Nature of Consciousness: Philosophical Debates*. Cambridge: MIT Press.

McLaughlin, B. P. 1992. “The Rise and Fall of British Emergentism,” in A. Beckermann,

H. Flohr, and J. Kim, *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. De Gruyter.

McLaughlin, B. P. 2001. “In Defense of New Wave Materialism: A Response to Horgan

and Tienson,” in C. Gillett and B. M. Loewer, *Physicalism and Its Discontents*.

Cambridge University Press.

McLaughlin, B. P. 2007. “On the Limits of A Priori Physicalism,” in B. P. McLaughlin

and J. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*. Blackwell.

Melnyk, A. 1997. “How to Keep the ‘Physical’ in Physicalism,” *Journal of Philosophy*,

94.12: 622-637.

Melnyk, A. 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge:

Cambridge University Press.

Merricks, T. 2001. *Objects and Persons*. Oxford: Oxford University Press.

Montero, B. 1999. “The Body Problem,” *Nous*, 33: 183-200.

Montero, B. 2001. “Post-Physicalism,” *Journal of Consciousness Studies*, 8: 61-80.

Montero, B. 2004. “Consciousness is Puzzling, But Not Paradoxical,” *Philosophy and*

*Phenomenological Research*, 69.1: 213-226.

Montero, B. 2006. “Physicalism in an Infinitely Decomposable World,” *Erkenntnis*, 64.2:

177-191.

Montero, B. 2012. “Review of *Physicalism* by Daniel Stoljar,” *Notre Dame Philosophical*

*Reviews*, URL = < http://ndpr.nd.edu/news/31706-physicalism/>.

Montero, B. and D. Papineau. 2005. “A Defense of the Via Negativa Argument for

Physicalism,” *Analysis*, 65: 233-237.

Nagel, E. 1961. *The Structure of Science*. New York: Harcourt, Brace.

Papineau, D. 1998. “Mind the Gap,” *Philosophical Perspectives*, 12: 373-389.

Papineau, D. 2002. *Thinking About Consciousness*. Oxford: Oxford University Press.

Perry, J. 2001. *Knowledge, Possibility, and Consciousness*. Cambridge: MIT Press.

Putnam, H. 1967. “Psychological Predicates,” in W. H. Capitan and D. D. Merrill (eds.)

*Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press.

Putnam, H. 1975. “The Meaning of Meaning,” in *Philosophical Papers, Volume II: Mind,*

*Language, and Reality*. Cambridge: Cambridge University Press.

Russell, B. 1921. *The Analysis of Matter*. George Allen and Unwin.

Shoemaker, S. 2001. “Realization and Mental Causation,” in C. Gillett and B. Loewer

(eds.), *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.

Shoemaker, S. 2007. *Physical Realization*. Oxford: Oxford University Press.

Shoemaker, S. 2013. “Physical Realization Without Preemption,” in S. C. Gibb, E. J.

Lowe, and R. D. Ingthorsson, *Mental Causation and Ontology.* Oxford; Oxford University Press.

Sider, T. 2003. “What’s So Bad About Overdetermination?” *Philosophy and*

*Phenomenological Research*, 67.3: 719-726.

Smart, J. J. C. 1978. “The Content of Physicalism,” *Philosophical Quarterly*, 28: 239-241.

Spurrett, D. and D. Papineau. 1999. “A Note on the Completeness of ‘Physics,’” *Analysis*,

59.1: 25-29.

Stoljar, D. 2001. “Two Conceptions of the Physical,” *Philosophy and Phenomenological*

*Research*, 62.2: 253-281.

Stoljar, D. 2009. “Physicalism,” *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.),

URL = <http://plato.stanford.edu/archives/fall2009/entries/physicalism/>.

Stoljar, D. 2010. *Physicalism*. Routledge.

Strawson, G. 2006. *Consciousness and its Place in Nature: Does Physicalism Entail*

*Panpsychism?* Imprint Academic.

Van Inwagen, P. 1990. *Material Beings*. Ithaca: Cornell University Press.

Van Inwagen, P. 1994. “Composition as Identity,” *Philosophical Perspectives*, 8: 207-220.

Wallace, M. 2009. *Composition as Identity.* Dissertation, University of North Carolina,

Chapel Hill.

Wallace, M. 2011a. “Composition as Identity: Part 1,” *Philosophy Compass*, 6.11: 804-816.

Wallace, M. 2011b. “Composition as Identity: Part 2,” *Philosophy Compass*, 6.11: 817-827.

Wilson, J. 1999. “How Superduper does a Physicalist Supervenience Need to Be?”

*Philosophical Quarterly*, 49: 33-52.

Wilson, J. 2005. “Supervenience-based Formulations of Physicalism,” *Nous*, 39.3: 426-459.

Wilson, J. 2006. “On Characterizing The Physical,” *Philosophical Studies*, 131: 61-99.

Wilson, J. 2009. “Determination, Realization and Mental Causation,” *Philosophical*

*Studies*, 145.1: 149-169.

Wilson, J. 2011. “Non-Reductive Realization and the Powers-Based Subset Strategy,” *The*

*Monist*, 94.1: 121-154.

Witmer, D. G. 2006. “How to Be a (Sort of) *A Priori* Physicalist,” *Philosophical Studies*, 131:

185-225.

Worley, S. 2006. “Physicalism and the Via Negativa,” *Philosophical Studies*, 131: 101-126.

Yi, B. 1999. “Is Mereology Ontologically Innocent?” *Philosophical Studies*, 93: 141-160.

1. Proponents include Spurrett and Papineau (1999); Levine (2001: 20); Montero (2001), (2004), and (2006); Papineau (2002: 41); Montero and Papineau (2005); and Worley (2006). [↑](#footnote-ref-1)
2. Montero (2001), Stoljar (2009: §11.4). I am unaware of anyone who has actually defended the simple version, but it is worth seeing the problem the position faces in order to motivate the more complicated versions that follow. [↑](#footnote-ref-2)
3. Wilson (2006: 61). Wilson defends a hybrid approach, combining elements of the via negativa with elements of the theory-based conception of ‘the physical’ that I reject in §3 below. I say “inter alia” in the text because Wilson’s NFM-constraint applies to physical entities generally (including substances, events, etc.), but here we will focus on properties. [↑](#footnote-ref-3)
4. See Lewis (1983), and (2009: 204), which explicitly equates fundamental properties with perfectly natural ones. Lewis unpacks the ‘nothing over and above’ claim in terms of supervenience: the fundamental properties serve as a minimal supervenience base for all other intrinsic properties. I do not follow him here; as I explain in §2, I reject supervenience-based formulations of ‘nothing over and above.’ [↑](#footnote-ref-4)
5. Kim (1996); Campbell (1997); Chalmers (1996); Montero (1999), (2001), (2004), and (2006). [↑](#footnote-ref-5)
6. Proponents include Lewis (1966), (1972), and (1994); Armstrong (1968); Jackson, Pargetter, and Prior (1982); and Jackson (1996). In response to Jackson (2002), Kim (2002) explains that his reductive view of the mind is consistent with realizer-functionalism but does not require it. [↑](#footnote-ref-6)
7. Lewis (1970), (1972). [↑](#footnote-ref-7)
8. In his early treatments—(1970) and (1972)—Lewis held that theoretical terms are denotationless at worlds where multiple *n*-tuples satisfy the given Ramsey sentence. But he later changed his mind about this. According to Lewis (1994: 58), the thing to conclude in such cases is that the denotation is “ambiguous.” This gives him a way to respond to the sort of multiple realization objection to the type identity theory famously advanced by Putnam (1967) and Fodor (1974). If, say, the pain-role is occupied at a given world by firing C-fibers in humans and by the distinct property of inflating D-tubes in Martians, then ‘pain’ ambiguously denotes both firing C-fibers and inflating D-tubes at that world, with context serving to resolve the ambiguity for particular utterances. For example, if you are engaged in conversation with fellow humans, context may determine that your utterance of ‘pain’ denotes firing C-fibers and not inflating D-tubes. For Lewis, theoretical terms are disguised definite descriptions (‘pain’ means *the* property occupying such-and-such role), and so their ambiguous denotation can be understood as a special case of the denotational ambiguity facing many such descriptions. Compare: sometimes ‘the car’ denotes my car, sometimes it denotes yours. [↑](#footnote-ref-8)
9. I have in mind the sort of causal exclusion reasoning set out most notably by Kim (1998). Lewis (1994) endorses such reasoning in the context of defending realizer-functionalism against other mind-body theories, including role-functionalism. Jackson (1996) takes a similar line. [↑](#footnote-ref-9)
10. On structural properties, see for instance Armstrong (1978: vol. II, pp. 68-71) and (1986). Lewis (1986) criticizes Armstrong’s account of structural *universals*, but allows that we can make sense of structural properties if properties are understood in other terms—for instance, as tropes. [↑](#footnote-ref-10)
11. I assume that neuroscience is multiply realized by fundamental physics. But realizer-functionalists should then treat such multiple realization just as they treat it elsewhere: splinter the property of firing C-fibers into a number of different properties, each of which can be identified with some complex structural property described by fundamental physics, and then take the term ‘firing C-fibers’ to be ambiguous in its denotation among these properties. See Lewis (1969) and (1994); Jackson, Pargetter and Prior (1982); Jackson (1996). [↑](#footnote-ref-11)
12. Cf. Merricks (2001), who uses causal exclusion reasoning to defend the view that there are no (inanimate) composite material objects—there are no baseballs, only atoms arranged baseballwise. The argument in the texts extends Merricks’s reasoning from composite material objects to structural properties. [↑](#footnote-ref-12)
13. Several philosophers have defended this response, but see especially Sider (2003), who deploys the overdeterminationist view against Merricks (2001). [↑](#footnote-ref-13)
14. Lewis (1994) and Kim (2005: 46-52) both argue that the overdeterminationist position is absurd in the mind-body case, and reject role-functionalism on the basis that it would require an overdeterminationist account of mental causation. [↑](#footnote-ref-14)
15. Different authors defend different versions of the principle. Lewis (1991: §3.6) holds that composition is closely analogous to identity (it is identity “in a loose and popular sense”). Baxter (1988a) and (1988b) holds that composition literally is identity, but then rejects the indiscernibility of identicals. Wallace (2009), (2011a), and (2011b) holds that composition literally is classical identity, and so obeys the indiscernibility of identicals. [↑](#footnote-ref-15)
16. Cf. Wallace (2009: Ch. 5) and (2011), who uses composition as identity in response to Merricks (2001). [↑](#footnote-ref-16)
17. Cf. Lewis (1991: 81), who in the context of defending his version of composition as identity writes, “If you draw up an inventory of Reality according to your scheme of things, it would be double counting to list the [parts] and then also list [the whole].” [↑](#footnote-ref-17)
18. Kim (2005: Ch. 2) takes up something like the challenge we have been considering, framing the point in terms of whether the causal exclusion problem generalizes, and whether all causal powers drain down to the microscopic level. Kim denies this, and explicitly rejects Merricks’s view, but his own positive position is less than fully clear. My recommendation to Kim is to embrace composition as identity. [↑](#footnote-ref-18)
19. Thanks to anonymous referees for forcing me to say more about the view just set out. [↑](#footnote-ref-19)
20. Critics of composition as identity include Van Inwagen (1994), Yi (1999), and Merricks (2001: 20-28). [↑](#footnote-ref-20)
21. The scenario might involve the brain being remarkably sensitive to its underlying microphysics, in order to allow instances of the given charge property to cause winces and groans. Kane (1998) proposes something like such sensitivity to explain how quantum indeterminacy might allow for libertarian free will, but for present purposes we can set the indeterminacy angle aside. While this sort of sensitivity to fundamental physics might be empirically unlikely, I deny that it is inconceivable. [↑](#footnote-ref-21)
22. See Lewis (1973) and (1983), and Armstrong (1983), who agree on such metaphysical contingency despite holding otherwise very different views about the laws of nature. [↑](#footnote-ref-22)
23. Standard realizer-functionalists deny that ‘pain’ is a rigid designator, taking mental terms to denote different properties in different worlds; see for instance Lewis (1994: 61-2). Consequently, realizer-functionalists will hold that pain is identical with the given charge property at *w* even though the identity does not hold at the actual world. Also, I note that I am assuming the charge property in question is intrinsic, in compliance with the view held by some that all perfectly natural properties are intrinsic. It is intrinsic, but the property qualifies as the denotation of ‘pain’ at *w* because of the relations it enters into there (‘pain’ is a kind of relational designator of an intrinsic property). Thanks to an anonymous referee on this point. [↑](#footnote-ref-23)
24. Chalmers (1996). [↑](#footnote-ref-24)
25. Functional definitions inevitably treat some terms as primitive—namely, the terms in the Ramsified theory that are not replaced by existentially bound variables. But which terms in the psychological theory used to generate a functional definition of ‘pain’ are to be taken as primitive? If I say *the physical terms*, my definition threatens to be viciously circular: in defining ‘the physical,’ I would be relying on a prior grasp of which term are the physical terms. Thankfully, the circularity can be avoided. The terms to be treated as primitive are the *non-mental terms* of the theory; that is, go through the theory and replace every mental term with an existentially bound variable, leaving all the other terms in place. This assumes a prior grasp on which terms are mental, but there is nothing circular about this for the purpose of defining ‘the physical.’ Thanks to an anonymous referee for pressing me on this point. [↑](#footnote-ref-25)
26. Chalmers (1996: 150-160) entertains epiphenomenalism but does not commit himself to it; Jackson (1982) commits himself. If, plausibly, you think that the case for why nonphysical qualia here in the actual world would be epiphenomenal depends in part on the actual laws of physics (e.g., the conservation laws), and you think that the laws of nature are metaphysically contingent, then you should allow that nonphysical qualia could be causally efficacious in possible worlds with different laws. [↑](#footnote-ref-26)
27. For supervenience-based formulations, see Davidson (1970), Lewis (1983), and Chalmers (1996). [↑](#footnote-ref-27)
28. See Horgan (1993), Kim (1998), Wilson (1999) and (2005), Melnyk (2003), and Shoemaker (2007). [↑](#footnote-ref-28)
29. Still, there is room for important disagreement. On my account, nonreductive physicalism collapses into role-functionalism, while reductive physicalism collapses into realizer-functionalism, and so there is room for physicalists to disagree about the status of reductionism. Thanks to an anonymous referee here. [↑](#footnote-ref-29)
30. In that case, the machinery of the Ramsey-Lewis method is not *needed* to specify behaviorist definitions—the great advantage of the method is that it can specify definitions for interdefined mental terms—but it still can be applied. [↑](#footnote-ref-30)
31. *Objection*: Wilson (1999) and (2011) defends a version of nonreductive physicalism that draws on the “subset account” of realization, which is at least not explicitly a form of role-functionalism. And Shoemaker (2001) explicitly rejects role-functionalism while defending his version of the subset account. *Reply*: Shoemaker (2013: 41) has recanted, and now says that the subset account is a version of the role-functionalist view. I agree with this later view of Shoemaker’s. *Objection*: Yablo (1992) defends a nonreductive physicalist view on which mental properties are determinables while physical properties are determinates, and he does not explicitly couch his position in role-functionalist terms. *Reply*: First, I agree with both Wilson (1999), (2009), (2011) and Shoemaker (2001) and (2007) that the determinate/determinable relation can be understood in terms of some version of the subset account of realization. Second, I once again agree with Shoemaker (2013: 41) that the subset account entails a role-functionalist view. Thanks to an anonymous referee here. [↑](#footnote-ref-31)
32. One general objection to via negativa approaches that I won’t be addressing is developed by Stoljar (2010: 87-88). Physicalism is intuitively inconsistent with emergent biological properties of the sort posited by vitalism, says the objection, and yet my definition entails that such properties qualify as physical, assuming they are neither fundamentally mental nor possibly realized by the fundamentally mental—they are instead fundamentally biological. This is an important objection but I am mostly satisfied with the response developed by other via negativa proponents, including Papineau (2002) and Montero (2012). Following their lead, I say that *if* there is a problem here it can be solved by amending my proposed definition so that a property is physical just in case it is neither fundamentally mental, *nor fundamentally biological*, nor possibly realized by such properties. Maybe further disjuncts will be needed as well—for example, disjuncts to rule out fundamentally moral properties (if these are taken to be distinct from mental properties), or perhaps fundamentally mathematical ones, or fundamentally metaphysical ones (e.g., Platonic forms). Still, there is no obvious reason to think that what needs to be added is hopelessly open-ended, ruining my proposal. At any rate, I set this problem aside in what follows, working with the unamended version of my proposal, focusing just on the mental. [↑](#footnote-ref-32)
33. Stoljar (2009: §10) and (2010: Ch. 4). Proponents include Smart (1978), Lewis (1994), Chalmers (1996), Melnyk (1997) and (2003), and Dowell (2006). [↑](#footnote-ref-33)
34. I do concede that for some philosophical purposes, the theory-based conception is appropriate. Philosophers of chemistry discuss whether chemical properties can be identified with physical properties. Their question is not whether chemical properties are fundamentally mental—of course they are not (at least barring panpsychism)—but whether such properties can be fully captured by physics. My claim in the text, then, is that for the *primary* role it plays in the philosophy of mind and metaphysics, the theory-based conception should be rejected and physicalism should not be equated with physicalismT. [↑](#footnote-ref-34)
35. Hempel (1969), (1980). [↑](#footnote-ref-35)
36. To relate this back to the issues raised in n. 33, I say that my definition of ‘the physical’ does not need to include a disjunct ruling out fundamentally chemical properties, because the instantiation of such properties is consistent with physicalism (properly understood). [↑](#footnote-ref-36)
37. Lewis (2009) explicitly endorses applying his general account of theoretical terms to the science of physics. There is a question of which terms in the physical theory should be left primitive through the process of Ramsification. One option, defended by Chalmers (2003) and (2012), is to treat as primitive logical, mathematical, nomic, and perhaps spatiotemporal terms; cf. Hawthorne (2001: 369-370). [↑](#footnote-ref-37)
38. Some philosophers, especially those drawing on Kripke (1980) and Putnam (1975), are skeptical of Lewis’s approach to defining theoretical terms. My own view is that the insights of Kripke and Putnam can be incorporated within Lewis’s framework, but for those philosophers who deny this I have a backup proposal. The argument that follows proceeds by considering a series of conceivable scenarios (i.e., scenarios that cannot be ruled out a priori) in which the comprehensive theory of physics is true. My argument should go through if you agree with my assignment of denotations to physical terms within these scenarios, even if you reject the machinery I use to make the assignment—that is, even if you reject my use of the Ramsey-Lewis method. The leading alternative accounts of the semantics of theoretical terms should agree with my assignments, at least within the scenarios considered. [↑](#footnote-ref-38)
39. A complication. You might hold that the functional definitions of the theoretical terms of physics should take the form of *rigidified* definite descriptions, so that ‘being an electron’ is defined as the property that *actually* occupies the electron role. In that case, assuming that some non-mental property occupies the electron-role here in the actual world, it follows that the term is denotationless at Idealism World. There are no electrons at Idealism World, there are merely instance of some other property that occupies the electron-role there. Compare: there is no water on Twin Earth, there is just some other substance (XYZ) that occupies the water-role there.

    If ‘being an electron’ is rigidified in this way, I can reframe the argument. Don’t think of Idealism World as an alternative possible world, think of it as a way the actual world might turn out to be. Would physicalism be true in that case (to anticipate the discussion that follows)?

    The sort of two-dimensionalist semantic framework developed by Chalmers (1996) and Jackson (1998) can be used to clarify things. If ‘being an electron’ is defined using a rigidified definite description, then the term will be denotationless at Idealism World when the world is considered as counterfactual, but it will denote the fundamentally mental property that occupies the electron-role there when the world is considered as actual. [↑](#footnote-ref-39)
40. Lewis (2009), Chalmers (1996: 134-136) and (forthcoming). See also the discussion in Stoljar (2001: 258-261). [↑](#footnote-ref-40)
41. Russell (1921). Chalmers (forthcoming) calls the position *Russellian panpsychism.* Strawson (2006) defends a broadly similar view, but explicitly rejects the theory-based conception in favor of what Stoljar (2009: §10) calls the object-based conception of ‘the physical.’ For a compelling objection to the object-based conception, see Montero (1999 and 2012). [↑](#footnote-ref-41)
42. Montero (1999: 191-193), drawing partly on Chomsky (1993 and 1995), develops a somewhat similar argument. But Montero frames the point in terms of the second horn of Hempel’s dilemma: perhaps some future physics will invoke fundamentally mental properties. Again, I claim my argument is independent of the dilemma. Suppose we were to grant that present physics is complete (even though this seems unlikely), thereby allowing defenders of the theory-based conception to avoid invoking future physics. Still, I claim, the theoretical terms of present physics might turn out to denote fundamentally mental properties, in which case physicalism (properly understood) would be false while physicalismT would be true. [↑](#footnote-ref-42)
43. Stoljar (2009: §11.3) advances this argument, drawing on Lewis (1983). Significantly, however, Lewis concedes that panpsychism is difficult to square with a functionalist view of mind, and so perhaps he would further concede that *if* physicalism requires functionalism—as I have argued—then physicalistic panpsychism is ruled out. [↑](#footnote-ref-43)
44. Wilson (2006: 79). Montero (1999) and (2001) takes a similar stance against physicalist panpsychism. [↑](#footnote-ref-44)
45. See for instance Hill (1991) and (1997), Loar (1997), Papineau (1998) and (2002), Block and Stalnaker (1999), McLaughlin (2001) and (2007), and Perry (2001). [↑](#footnote-ref-45)
46. This formulation of the argument is taken from Chalmers (forthcoming). [↑](#footnote-ref-46)
47. Here, we focus on a priori *type* physicalism, as opposed to a priori physicalism more broadly, to align the discussion with the paper’s focus on the type identity theory. On the relevant sense of entailment, one class of truths entails another just in case in any logically possible world in which the first class of truths obtains, the second class of truths obtains. An entailment is said to be a priori if it is knowable a priori. It is common in discussions of this sort of view to add indexical truths (e.g., truths using terms like ‘I’ and ‘now’) and negative truths (e.g., ‘there are no angels’) to the entailment base, but we will ignore this complication here. [↑](#footnote-ref-47)
48. I myself agree with realizer-functionalists that the zombie hypothesis is inconceivable. I will not attempt to defend this position here, other than to note my sympathy with the view defended by Jackson (2003b). [↑](#footnote-ref-48)
49. Such philosophers accept that *P* entails *M*, in that there is no logically possible world where *P* obtains but *M* does not, but they deny that this entailment is knowable a priori. [↑](#footnote-ref-49)
50. Or, more often, they provide an account of our mental concepts, and especially our phenomenal concepts. Here, I focus on language to connect the view more easily to the realizer-functionalist framework. [↑](#footnote-ref-50)
51. Cf. Chalmers (1996), Jackson (1998, 2003, and 2003b), Chalmers and Jackson (2001), Kim (2005), Witmer (2006). [↑](#footnote-ref-51)
52. Kim (1998: Ch. 4) and (2005: Ch. 4). Kim’s account is meant to improve on Ernest Nagel’s (1961) view. [↑](#footnote-ref-52)
53. In Kim’s (1998: 98-99) initial presentation of his model, he endorses such property identifications. Later, in response to the problem posed by multiple realization, he considers (but does not commit himself to) alternatives, including eliminating the property described by the reduced theory, or identifying it with the disjunction of the different realizer properties. See Kim (1998: 106-112) and (2002). [↑](#footnote-ref-53)
54. Kim (1998: 98). [↑](#footnote-ref-54)
55. On my view theoretical terms generally are functionally definable, and so the asymmetry is not that the terms of the reduced theory have functional definitions while those of the reducing theory do not. Rather, it’s that the terms of the reduced theory can be functionally defined using the vocabulary of the reducing theory, but not vice versa. If psychology is ultimately reducible to physics, then ‘pain’ can be defined in terms taken from physics, while ‘being an electron’ cannot similarly be defined in psychological terms. Cf. Chalmers and Jackson (2001). [↑](#footnote-ref-55)
56. See for example Horgan and Tienson (2001), Chalmers and Jackson (2001), Kim (2005: Ch. 5), Jackson (2007), and Horgan (2010). [↑](#footnote-ref-56)
57. See for example Block and Stalnaker (1999), McLaughlin (2001) and (2007). [↑](#footnote-ref-57)
58. That is, an argument whose first premise is that it is conceivable that all the truths described by physics are just as they are but that Chalmers lacks neural states, and whose conclusion is that neural states are something over and above those states described by physics. Block and Stalnaker (1999) raise wide-ranging doubts about functional definitions generally; I accept much of the reply in Chalmers and Jackson (2001). [↑](#footnote-ref-58)
59. Cf. Montero (2001: §5). [↑](#footnote-ref-59)
60. The *proper-part-of* relation that obtains between a whole and any one of its proper parts is asymmetric, but this relation cannot do the work needed here, for a whole clearly is ‘something over and above’ any particular proper part—if for no other reason, then because it has other proper parts. What matters is the relation between a whole and *all* of its parts. If composition is identity, then a whole indeed is nothing over and above all its parts. But in that case, the problems mentioned in the text arise. [↑](#footnote-ref-60)
61. Merricks (2001) cites the British Emergentists as an inspiration, while distinguishing his argument from theirs. On the British Emergentists, see McLaughlin (1992). [↑](#footnote-ref-61)
62. Although, again, I cannot claim to have covered all the options. Jenkins (2011) explores views of metaphysical dependence that would allow that pain could be both identical with firing C-fibers and dependent on firing C-fibers (in a way that firing C-fibers are not similarly dependent on pain). I am skeptical, but won’t try to justify my skepticism with an argument here. [↑](#footnote-ref-62)
63. Many thanks to Zachary Barnett, Tomas Bogardus, Janice Dowell, Steven James, Barbara Montero, Bryan Pickel, Ariela Tubert, and Gene Witmer for discussion. Thanks as well to an audience at the 2015 Pacific APA for feedback. Finally, thanks to the anonymous referees who provided me with excellent, patient comments. [↑](#footnote-ref-63)