

CHAPTER I

*Supervenience and Nonreductive Physicalism***1.1 Introduction**

In the Introduction, I described the levels-based physicalism that has come to dominate the philosophical landscape. In the next several chapters, I consider a variety of strategies for articulating nonreductive, levels-based physicalism. My conclusion is that none succeed. While my focus is on physicalist metaphysics, the rejection of levels of reality in physicalism may generalize to any attempt to understand reality as having the kind of levels-based structure exemplified by nonreductive physicalism.

My critique of nonreductive physicalism begins by considering the characterization of physicalism as the thesis that all of reality *supervenes on* how things are physically. On this approach to physicalist metaphysics, physicalism regarding the mind consists in the claim that for any thought, experience, or emotion that one might have, there is some physical state of the world upon which that mental state depends and that guarantees the occurrence of that mental state. There are several reasons to begin this way. Supervenience has played a central role in metaphysics since around 1970. Nonreductive physicalism has frequently been conceived in terms of supervenience, and while supervenience-based definitions of physicalism are now often dismissed without argument, it is not difficult to understand why supervenience might have seemed, in Kim's words, "tailor-made" for understanding "microphysical determination without microphysical reducibility" (1978, 151). The perceived shortcomings of supervenience-based definitions have also motivated alternative accounts of physicalism, especially the realization- and Grounding-based accounts that I consider in later chapters. Moreover, despite current hostility toward the philosophical usefulness of supervenience, a number of influential criticisms of supervenience-based physicalism are less straightforward than has often been assumed – contrary to current fashion, it is not *obvious* that sufficient conditions for physicalism

cannot be stated in terms of supervenience. More importantly, considering why this is the case is instructive, and in the end dashes any hope that supervenience might be the key to vindicating nonreductive physicalism.

I will begin by sketching the notion of supervenience, its application to physicalist metaphysics, and some reasons why it has been thought that supervenience cannot give sufficient conditions for physicalism.

1.2 The Rise and Fall of the Supervenience Approach to Physicalism

The use of “supervenience” in the philosophy of mind is usually traced to Donald Davidson’s claim, in his 1970 “Mental Events,” that the denial of strict psychophysical laws “is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics,” such that “there cannot be two events alike in all physical respects but differing in some mental respect” (1970/2001, 214).¹ While the concept of supervenience has since been articulated in many ways, the core idea that it is intended to express is that of certain items being strictly necessitated or determined by others.² Consider the following “supervenience thesis”:

S1. Necessarily, if some property M is instantiated at some time t, some physical property P is instantiated at t such that necessarily, given the physical laws, if P is instantiated, then so is M.

S1 is intended to express a kind of synchronic, noncausal dependence and determination: *synchronic*, because the physical property is taken to be instantiated at the very same time or times that the supervenient property is instantiated; *noncausal*, because aside from being synchronic, the relation between physical occurrences and instances of “supervenient properties” need not involve features often associated with causation, such as the transfer of energy. So, if S1 is true, whenever someone has an experience, there must be some physical state of the world, at that very time, that together with the physical laws guarantees the occurrence of that

¹ For this account of the introduction of “supervenience” into the philosophy of mind, see Bennett and McLaughlin (2014). The history of “supervenience” has been documented elsewhere; see Bennett and McLaughlin (2014), Horgan (1993), and Kim (1990). There are many questions that could be raised about Davidson’s claim that the absence of psychophysical laws is consistent with mind–body supervenience; see Davidson (1993), Kim (1984, 1993), and McLaughlin (1993).

² For discussion of the various ways of defining supervenience, see Bennett and McLaughlin (2014), Chalmers (1996), Horgan (1982, 1993), Howell (2009), Jackson (1998), Kim (1978, 1984, 1987), Lewis (1983), and McLaughlin (1995).

experience. Generalizing, the idea is that the entire distribution of properties in the world is strictly fixed by the physical character of the world.

There are many issues that could be raised about S₁, and S₁ could be refined in many ways. However, it is widely agreed that for S₁ to have any hope of providing sufficient conditions for physicalism, it must be a thesis of “metaphysical” supervenience; this amounts to taking the second necessity operator in S₁ to express metaphysical necessity, as claiming that in all possible worlds with the same physical laws, the instantiation of P is accompanied by the instantiation of M. For S₁ to have any hope of defining a physicalist metaphysic, the second necessity operator must mean *really* necessary! The intended contrast is with weaker grades of modal dependence. Suppose that the second necessity operator in S₁ is taken to merely express “nomological” necessity, that is, necessity as a matter of natural law. In this case, S₁ says that whenever some property M is instantiated, some physical property P must be instantiated such that given the physical laws, M must be instantiated as a matter of natural law. So understood, S₁ does not entail that the distribution of mental properties is strictly determined by the physical character of the world; it only entails that the distribution of mental properties is strictly determined by the physical character of the world in conjunction with “psychophysical laws” correlating physical properties with mental properties. Why is this not enough for physicalism? The reason is that patently dualistic views, perhaps even a Cartesian dualism of mind and body, can accept lawful correlations between mental properties and physical properties.³ Generally, there is little reason to think that mere nomic correlation entails that one of the correlated items is nothing over and above the other. In terms of a familiar but useful metaphor, physicalism should suppose that all God had to do to create the entire world is to create the physical world – distribute the physical properties and fix the physical laws that govern them.⁴ Taking the second necessity operator in S₁ to express metaphysical necessity has this result. However, taking it to express nomological necessity does not, as this allows that in order to create the entire world, God had to distribute the physical properties, specify the physical laws governing the physical properties, and

³ Some have thought that a Cartesian view can take the mental properties of mental substances to metaphysically supervene on physical properties, and on this basis have denied that S₁ can suffice for physicalism even if the second necessity operator is taken to express metaphysical necessity; see Section 1.4.

⁴ As Kripke (1972/1980) suggests.

then moreover specify how physical properties correlate with, say, mental properties.⁵ Anyone attracted to supervenience-based physicalism should thus suppose that “upgrading” the second necessity operator in S₁ to express metaphysical necessity secures a more intimate relation between the physical and anything other than the purely physical.

Related remarks apply to a “global supervenience” thesis such as the following:

S₂. All worlds that are physically just like the actual with no additional stuff (all “minimal physical duplicates”) are like the actual world in all respects (“duplicates simpliciter”).

Taking S₂ to express metaphysical supervenience amounts to taking the initial quantifier to range over metaphysically possible worlds as opposed to some more restricted set. S₂ then says that duplicating how the world is physically strictly suffices to duplicate how the world is generally. So understood, S₂ entails that all worlds that are physically the same as the actual world with nothing “extra” beyond this are exactly the same with respect to how thought and consciousness is distributed. As with S₁, the intended contrast is with mere nomological dependence.

On supervenience-based physicalism, physicalism is true just in case a thesis like S₁ or S₂ obtains. Why might this seem promising as a means for characterizing an outlook that is physicalist and nonreductive? The physicalist component has already been sketched – it is the idea that fixing the physical nature of the world suffices to fix everything about the world. Regarding the nonreductive component, the key idea is simply that nothing in theses such as S₁ and S₂ seems to require the reducibility of supervenient properties or instances of them. For example, S₁ seems consistent with supposing that consciousness is distinct from anything physical, even if the second necessity operator is taken to express metaphysical necessity. This is especially so if S₁ and S₂ are taken to involve a picture akin to Figure 1.1.

Given this, it is not surprising that supervenience has sometimes seemed ideally suited to characterize a nonreductive physicalist position regarding mind and world. Nonetheless, there are several well-known

⁵ Kim (1998) notes the compatibility of nomological supervenience with an array of views in the metaphysics of mind. For discussion of the distinction between nomological supervenience and metaphysical supervenience, see Bennett and McLaughlin (2014), Chalmers (1996), Howell (2009), and Van Cleve (1990). I have omitted discussion of how the first necessity operator in S₁ should be interpreted. The issue here concerns the extent to which physicalism should suppose that the properties that are instantiated in the actual world must have physical bases. While I regard this issue as both unsettled and potentially significant, I will continue to set it aside; for discussion, see Jackson (1998), Lewis (1983), and Levine and Trogon (2009).

Supervenience and Nonreductive Physicalism

25

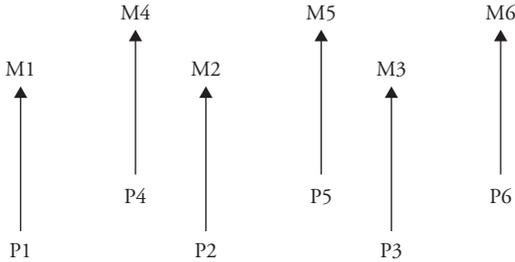


Figure 1.1 Supervenience and levels.

The Ms represent occurrences of supervenient properties, while the Ps are physical states of the world upon which the Ms depend and which determine or necessitate the Ms. This representation comports with the idea that S1 and S2 provide no basis for thinking that the Ms are reducible to the Ps. Rather, aside from being connected to the Ps by the vertical arrows, the Ms appear wholly distinct from the Ps.

reasons to question whether supervenience-based physicalism can deliver this happy result. The most important and interesting objections are those that turn on the claim that nonphysicalist positions are consistent with theses such as S1 and S2. Stoljar (2010) calls this the “objection from necessitation dualism” to supervenience-based definitions. One reason to think that nonphysicalist positions *can* endorse S1 and S2 is that nonphysicalist positions *have* endorsed S1 and S2! In his influential development of this objection, Horgan (1993, 2006) appeals to the emergentism of Broad (1925) and others in the first part of the twentieth century, as well as the metaethical view often attributed to G.E. Moore (1903). As Horgan reads him, Broad held that while all properties that are instantiated are strictly determined by physical properties, in some cases there is no explanation for why this is so, and that in these cases the properties so determined are genuinely novel with respect to physical properties. Broad seems to have held that such properties – the “emergent” ones – have distinctive, irreducible causal powers with respect to the physical domain. If this is right, Broad was committed to denying the causal completeness of the physical domain, that every physical event that has a sufficient cause at a time has a sufficient physical cause at that time.⁶ Horgan takes

⁶ For related work in the “British emergentist” movement, see Alexander (1920) and Morgan (1923). For useful discussion, see Kim (1999) and McLaughlin (1992). It can be questioned whether Broad, or any British emergentist, held that emergent properties arise from physical properties as a matter of metaphysical rather than mere nomological fact. In response, Horgan contends that even if they did not take emergent properties to metaphysically supervene on physical properties, it would have been consistent for them to do so (1993, 559–60; 2006, 160). For remarks in this

Moore to have similarly held that goodness is a simple and nonnatural property, that whether something is good is strictly determined by what “natural” properties are instantiated, and that this determination of goodness by natural properties does not admit of explanation.⁷ Horgan’s point is that neither Broad’s emergentism nor Moore’s metaethics are consistent with a comprehensive physicalism; but as they seem consistent with, and indeed to positively endorse, strong supervenience theses, no thesis along the lines of S1 or S2 can suffice for physicalism.

Why are these outlooks inconsistent with physicalism? Horgan identifies two points of concern, both implicit in the preceding sketch:

- (1) Brute supervenience: brute, inexplicable relations of supervenience, “*sui generis* principles of metaphysical necessitation” (2006, 160)
- (2) Supervenient dualism (“dualistic” or “nonphysicalist” novelty): properties or features that have no place in a physicalist outlook supervening on physical properties.

According to Horgan and others,⁸ (1) and (2) are antithetical to physicalist metaphysics; so, emergentism and Moorean metaethics are counterexamples to supervenience-based definitions of physicalism because they endorse at least one of (1) and (2) and yet also endorse S1 and S2. In terms of the distinction between nomological and metaphysical supervenience, the charge is that metaphysical supervenience fails to rule out the dualism that nomological supervenience allows.

The dialectical progression just traced represents conventional wisdom regarding why supervenience-based definitions might seem attractive and also why they prove unsatisfactory. Following some recent discussions,⁹ however, my view is that the supervenience-based approach cannot be so blithely dismissed. To begin: Why should it be thought, as Horgan contends, that brute metaphysical supervenience, “*sui generis* principles of metaphysical necessitation,” has no place in physicalist metaphysics? Similarly, why should it be thought that physicalism requires that S1 and S2

direction, see also Wilson (2002, 2005) and Vision (2011). If this is correct, emergentism cannot be contrasted with physicalism in terms of a commitment to nomological, rather than metaphysical, supervenience, as in Van Cleve (1990). For doubts about whether emergentism *is* consistent with metaphysical supervenience, see Kim (1999, 2011) and Howell (2009); see Sections 1.4 and 1.5 for discussion.

⁷ There are nontrivial issues about just what Moore meant in calling goodness “nonnatural”; see Dreier (2006) and Polger (2013).

⁸ In her critique of supervenience-based physicalism, Wilson (1999, 2005) focuses largely on (2) and downgrades the significance of (1); for a related position, see Polger (2013). Melnyk (2003, 2006), in contrast, emphasizes (1). See Morris (2018) for discussion of Wilson on brute supervenience; see Morris (2018) and Section 1.3.1 for discussion of Polger on physicalism and brute supervenience.

⁹ See Howell (2009), Kim (2011), and Polger (2013).

be explainable – that physicalism requires what Horgan calls “superdupervenience,” “ontological supervenience that is robustly explainable in a materialistically acceptable way” (1993, 577)?

1.3 What is Wrong with Brute Supervenience?

Despite the prevalence of the claim that no physicalism can accept “inter-level principles expressing brute, fundamental, metaphysical-necessitation relations” in his widely cited critique of supervenience-based physicalism, Horgan says surprisingly little in defense (2006, 160). Nor is it obvious that this is true. For instance, one might think that *if* properties that are strictly necessitated by physical properties are all physically acceptable, it is neither here nor there whether this necessitation *itself* admits of explanation. Moreover, even if neither emergentism nor Moorean metaethics are consistent with physicalism and they accept brute supervenience, it does not follow that brute supervenience is itself physically unacceptable. After all, as Horgan recognizes, that neither emergentism nor Moorean metaethics are physicalist positions can be accounted for in terms of a commitment to dualistic properties supervening on physical properties.¹⁰

One might think that no physicalist position can accept brute inter-level supervenience because it is closely connected to the possibility of properties that have no place in a physicalist framework supervening on physical properties. However, as I argue in Section 1.4, the issues surrounding whether dualistic properties may metaphysically supervene on physical properties are themselves less than transparent. Moreover, this is *not* the strategy that Horgan pursues. He rather holds that brute supervenience has no place in physicalism independent of any connection between brute supervenience and dualistic properties supervening on physical properties. This is the import of his remark that emergentism would not be a form of physicalism “even if it did affirm the causal completeness of physics,” given its commitment to “supervenient properties whose supervenience is not materialistically explainable” (1993, 560).

1.3.1 *Brute Supervenience and Unexplained Explainers*

The most substantive argument that Horgan offers for thinking that there is something about brute supervenience as such that conflicts with physicalist scruples turns on a claim about the role of “unexplained explainers” in physicalist metaphysics. The claim is that physicalism should hold

¹⁰ Polger (2013) makes a similar point.

that “any metaphysically basic facts or laws – any unexplained explainers, so to speak – are facts or laws within physics itself” (1993, 560). The reason that no physicalism can accept brute supervenience, then, is that this would entail “unexplained explainers” outside of physics, “metaphysically basic facts” beyond the “facts or laws within physics itself.” Straightforwardly, if there is no explanation for why S_1 or S_2 holds, such theses might themselves be regarded as a “metaphysically basic facts” or “unexplained explainers” not “within physics itself.” When cast in this way, Horgan’s claim that physicalism requires supervenience that is explainable in a materialistically acceptable way amounts to requiring for the truth of physicalism that S_1 and S_2 be explainable without the use of unexplained explainers from outside of physics.

There is something intuitive about the claim that on a comprehensive physicalism, all basic facts are within physics and that S_1 and S_2 should be explainable without any unexplained explainers from outside of physics. Perhaps it is possible to define “unexplained explainer,” “metaphysically basic,” “fact,” and “physics” so that there is a basis for holding that no physicalism can include brute supervenience. However, accomplishing this is not an easy task, and I am not sure that there is a wholly convincing reason to think that brute supervenience as such is antithetical to physicalism. There are somewhat different reasons, however, for questioning whether physicalism requires that supervenience theses be explainable without unexplained explainers from outside of physics, which in turn yield different perspectives on Horgan’s insistence that physicalism needs materialistically explainable supervenience.

In his ambitious critique of Horgan on physicalism and supervenience, Polger (2013) argues that Horgan’s ban on brute inter-level supervenience and his claim that physicalism requires that all unexplained explainers are within physics should be rejected because they lead to an absurd regress.¹¹ Polger begins by diagnosing “physicalism’s alleged incompatibility with brute, inter-level metaphysical necessity” as stemming from the thought that inter-level “determination relations must themselves be

¹¹ Polger draws from a similar line of thought due to Michael Lynch and Joshua Glasgow (2003). As Polger notes, the issues here are similar to those in recent discussions of metaphysical Grounding, see Bennett (2011a), Dasgupta (2015), and deRosset (2013); see also Sider (2011). I discuss Grounding-based physicalism in Chapter 5. The present discussion is specifically geared toward Polger’s use of this regress-theoretic concern as a response to Horgan’s rejection of brute supervenience in physicalism, and I do not claim that the considerations that I offer directly bear on related topics in the Grounding literature (such as the vexed question of what Grounds Grounding, and the pun-like elaborations on this question).

either physical or determined by the physical.” He thus takes the alleged physical unacceptability of brute supervenience to be rooted in the worry that such “supervenience relations are neither [physical or determined by the physical]” (2013, 83). If I have understood Polger correctly, the *problem* with holding that brute supervenience is antithetical to physicalism on this basis is that the very demand that “determination relations must themselves be either physical or determined by the physical” leads to the aforementioned regress, the argument for which proceeds as follows. Schematically, suppose that physicalism is defined as the view that everything is physical or stands in some relation R to how things are physically – for example, that everything is physical or supervenes on, and so is necessitated or determined by, how things are physically. It seems to follow that any relation R must itself either be physical or stand in R to how things are physically. Neither seems sustainable, however. First, suppose that R is not among how things are physically. If R does not stand in R to how things are physically, physicalism is false if it requires that all of reality, R-relations included, is either physical or stands in R to how things are physically. Yet if R does stand in R to how things are physically, this appears to lead to a regress: If R is required to stand in R to how things are physically, by parity of reasoning the R-relation between the original R-relation and how things are physically must stand in a further R-relation to how things are physically; by parity of reasoning, the further R-relation must itself stand in R to how things are physically, and so on. Second, R cannot be included in the physical base, as R is an inter-level relation, a relation *between* how things are physically and how things are generally. So, there is no place for R itself in a physicalism characterized as the thesis that everything is physical or stands in R to how things are physically.

Polger takes the concern to apply to any metaphysics that involves inter-level determination relations, with supervenience-based physicalism as an instance. That is:

Any theory that claims that everything is an A or determined by and dependent on the A's will face the same question about the determination and dependence relations, leading to the same regress ... Whatever is wrong here is not peculiar to physicalism. (2013, 84)

Polger's conclusion is that the “purported pretensions of physicalism to be completely comprehensive” should be reconsidered (2013, 85). Specifically, inter-level relations – including supervenience and necessitation relations or theses expressing such relations – should not themselves be included in the scope of physicalism. For example, physicalism should

require that consciousness is strictly determined by how things are physically, but that consciousness is determined by how things are physically need not itself be determined by how things are physically. This entails denying that physicalism requires that all unexplained explainers are within physics, as it supposes that inter-level relations, or theses expressing such relations, may be unexplained explainers outside of physics. Hence, in contrast to Horgan, physicalism should not require that “any metaphysically basic facts or laws ... are facts or laws within physics,” because physicalism may permit for inter-level relations, including those expressed by S_1 and S_2 , that are basic yet not within physics. It cannot be claimed that physicalism is incompatible with brute inter-level relations on the grounds that such relations are neither physical nor stand in any such relation to the physical, as the demand that they stand in such a relation yields a regress.¹²

Polger concludes that Horgan’s ban on unexplained explainers outside of physics is too strong, and for reasons that emerge below, I think that he may be right about this. However, I doubt that Polger’s argument wholly undermines Horgan’s position on physicalism and brute supervenience.

Suppose Polger is correct that any levels-based metaphysics must accept some unexplained explainers outside of physics. Even if this is conceded, S_1 and S_2 need not be among these unexplained explainers. Given this, even if any levels-based physicalism must accept some bruteness outside of the physical, it does not follow that any such framework may, or must, accept brute inter-level supervenience. Take, for example, the proposal that S_1 and S_2 should be explainable using some combination of physics, logic, math, conceptual truths, and a posteriori identifications of higher-level properties with physical properties and physically realized functional properties. This appears to be Melnyk’s view, as well as the view of many

¹² Despite favoring a type identity theory (see Polger [2004] and Polger and Shapiro [2016]), Polger does not seriously consider that the lesson from the alleged regress is simply that reality is not ordered by some such inter-level relation. This is curious, given that Lynch and Glasgow (2003) concede that the reasoning here only has force against nonreductive physicalism. Related remarks apply to certain discussions of Grounding. Shamik Dasgupta (2015), for example, supposes that regress-theoretic concerns about what Grounds Grounding jeopardize the very possibility of physicalism and on this basis develops substantive machinery designed to set aside this perceived threat to “the possibility of physicalism.” But one might alternatively conclude that physicalism is not profitably understood in terms of a distinction between the fundamental and the derivative, with the latter Grounded in the former. This is all symptomatic of the assumption that levels-based physicalism is just about inevitable; see the Introduction and Chapter 6.