13

Naturalism and Physicalism

BARBARA GAIL MONTERO AND
DAVID PAPINEAU

Introduction

Naturalism means different things to different people. But one significant strand in contemporary understandings of the term is *physicalism*. This is the doctrine that everything is physical. In this chapter, we shall examine this doctrine and assess the strength of the case in its favor.

Physicalism has increased markedly in popularity in the Western world over the past century or so. In a recent survey of philosophers, 56% of the 3000-plus respondents were in favor of physicalism, and only 27% definitely against.[1] This is a relatively new phenomenon. The proportions would have been very different in the 19th century. One issue to be addressed in this chapter is the explanation of this shift.[2]

As understood by contemporary philosophers, physicalism is a relatively laissez-faire doctrine. The basic thought is that everything is physically constituted. But this is generally understood in such a way as to avoid any strong methodological implications. Few contemporary physicalists would argue that the truth of physicalism means that we should close down the psychology, biology, or even meteorology departments and hand everything over to the physicists. The practicalities of studying complex structures like minds, bodies, and weather systems call for special methods and techniques beyond those used in the analysis of basic physical processes. Moreover, many physicalists would add that there is a metaphysical basis for this methodological precept. In their view, the properties and patterns that are displayed in the "special sciences" (psychology, biology, meteorology, etc.) are genuinely novel, in that they cannot be defined within the language of basic physics or explained by basic physical principles.

1 http://philpapers.org/surveys/results.pl (last accessed July 15, 2015).

2 For a detailed account of how attitudes toward physicalism in the Western world have changed over the past four centuries, see Papineau (2001).

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.

We shall return to some of these niceties further on. For the moment, it will suffice to characterize physicalism graphically. Imagine God creating the world, and in particular imagine that God's first task is to put all the physical material in place. Now we can ask: Given that God has arranged all the quarks, leptons, and so on, is it time to rest? Antiphysicalists will say "no," but physicalists will answer "yes." The antiphysicalists will feel that God still needs to add all the conscious thoughts and feelings. But physicalists will think that this has already been taken care of. By fixing the basic physical facts, God has therewith fixed all the facts, including conscious mental facts. Nothing more is needed for conscious minds than the relevant physical facts. Or, to put it the other way round, not even God could create a world physically just like ours but lacking conscious minds.[3]

This metaphor captures the technical notion of physicalism as the "metaphysically necessary supervenience" of all facts on the basic physical facts. This is the idea that everything is fixed once the fundamental physical facts are fixed: people and plants (and other such things), even though they are not necessarily reducible to quarks and leptons (and other such things), are in some sense nothing over and above quarks and leptons (and other such things). We shall understand "physicalism" in this sense in what follows. In line with this, we shall count some property or entity as "physical" not only when it is a basic physical property or entity, but also when it supervenes on the basic physical facts.

The Case for Physicalism

Despite the previously stated qualifications, physicalism is still a very strong doctrine, and indeed one that has been denied throughout much of history. Why do so many contemporary philosophers embrace it? As we see it, the driving motivation behind the commitment to physicalism is the need to explain how things that are apparently not physical can have physical effects. Thus, many contemporary thinkers adopt a physicalist view of the mental realm because they think that otherwise we would be unable to explain how mental processes can causally influence the physical world. Similar considerations motivate physicalist views of the biological and other realms.

It may not be immediately obvious why this need to account for physical influence should demand that we view the mental, biological, and other realms as themselves physical. After all, there seems nothing *a priori* incoherent in the idea of nonphysical agents exerting a causal influence on physical processes, as is testified by the conceptual cogency of traditional stories in which active spirits and other immaterial agents intervene in the physical world.

However, there may be *a posteriori* objections to such nonphysical interventions, even if there are no *a priori* objections. We shall see further on how modern scientific theory places strong restrictions on the kinds of entities that can have physical effects. Given that mental and biological phenomena clearly do have such effects, this suggests that they must satisfy the relevant restrictions.

We can put the argument like this. Science shows us that physical effects can always be accounted for by fully *physical* causes. But we know that biological and mental facts are

3 This metaphor is inspired by Kripke's discussion of the mind–brain identity theory in *Naming and Necessity* (1980, 153–154). Though most accept this account of physicalism, see Montero (2013) for an argument that physicalism is consistent with the denial of such a view.

often among the causes of physical effects (as when a dog's *breathing* reduces *the oxygen level,* or when I *decide to move that stone* and thereby *move it*). So those biological and mental causes must themselves be fully physical.

In effect, the scientific discovery that physical effects always have physical causes squeezes any *non*physical factors out of the realm of things that can affect the physical world. So, in order to account for how mental and biological processes do affect the physical world, we have to recognize that they are themselves physical.

Physicalists do not deny that there are mental and biological facts. Of course there are. They don't want to eliminate these facts, but illuminate their nature. In this spirit, they view mental and biological facts as physical. When we talk about *decisions* or *breathing*, we aren't talking about some extra facts, distinct from electrochemical goings-on in brains and bodies. Rather, we are just talking about large-scale aspects of the underlying electrochemical processes.

Note how the science-based argument outlined in this section (if mental and biological causes weren't physical, they would be squeezed out of the range of things that can affect the physical world) is different from some of the more traditional arguments against Cartesian dualism and similar nonphysicalist philosophies. Some of the earliest commentators on Descartes argued that he had divided mind and body too sharply to allow any causal interaction between them. It is not clear how telling this worry is. On many conceptions of causation there is no reason why there should not be causal intercourse between Descartes' two realms, and historically, as we shall see, it seems unlikely that this traditional concern did much to discredit Cartesian interactionist dualism. In any case, the science-based argument on which we shall focus is different. Its thrust is not that immaterial minds and other nonphysical entities are *a priori* the wrong kind of thing to affect the physical world, but simply that science has shown us *a posteriori* that the physical world is not subject to such influences.

Note also how the science-based argument only indicates that the types of things that have physical effects are themselves also physical; it says nothing about types of things that have no physical effects. For example, many philosophers think that there are abstract numbers and sets which inhabit some realm outside space and time and so are causally inert and lack any physical effects. Our science-based argument leaves it open whether such things must be physical. Accordingly, we shall understand "physicalism" in what follows not as the thesis that absolutely *everything* is physical (as we put it at the start), but as the more qualified claim that everything within the spatial or temporal realm is physical.

Leibniz, Newton, and the Conservation of Energy

It will be worth explaining in some detail the evolution of modern scientific ideas about the range of things that can have physical effects. This will help to forestall the impression that contemporary physicalism is some kind of fad. Thus, it is sometimes suggested that physicalism rests, not on reasoned argument, but on some kind of unargued commitment, some ultimate decision to nail one's colors to the physicalist mast. And this diagnosis can seem to be supported by the fact that physicalist doctrines have become widely popular in the Western world only in the past few decades. However, familiarity with the relevant scientific history casts the matter in a different light. It turns out that physicalist doctrines,

far from varying with ephemeral fashion, are closely responsive to received scientific opinion about the range of causes that can have physical effects.

Let us begin with the "mechanical philosophers" of the 17th century, who held that any material body maintains a constant velocity unless acted on, and moreover that all action is due to the impact between one material particle and another. So stated, the mechanical philosophy immediately precludes anything except impacting material particles from producing physical effects. Leibniz saw this clearly, and concluded that it discredited Descartes' interactive dualism, which had a nonmaterial mind influencing the physical world (Woolhouse 1985). (As it happens, Leibniz did not therewith reject dualism, but instead opted for "preestablished harmony." Views which avoid physicalist views of the mind by denying its causal efficacy will be discussed further on.)

At the end of the 17th century, Newtonian physics replaced the mechanical philosophy of Descartes and Leibniz. This reinstated the possibility of interactive dualism, since it allowed that disembodied nonimpact forces could cause physical effects. Newtonian physics was quite open-ended about the kinds of forces that exist. Early Newtonians posited distinctive mental and vital forces alongside magnetic, chemical, gravitational, and impact forces. Accordingly, they took fundamental mental action and fundamental vital action in the material world to be perfectly consistent with the principles of physics. Moreover, there is nothing in the original principles of Newtonian mechanics to stop fundamental mental forces arising autonomously and unpredictably, in line with common assumptions about the operation of the mind (Papineau 2001). As a result, the Newtonian world view was effectively an interactive pluralism that recognized a wide range of nonphysical influences, including spontaneous mental influences (or "determinations of the soul," as they would then have been called).

In the middle of the 19th century, the conservation of kinetic plus potential energy came to be accepted as a basic principle of physics (Elkana 1974). In itself, this does not rule out fundamental mental or vital forces, for there is no reason why such forces should not themselves be "conservative," operating in such a way as to compensate losses of kinetic energy by gains in potential energy and vice versa.[4] (The term "nervous energy" is a relic of the widespread late 19th-century assumption that mental processes store up a species of potential energy, which action then converts into the kinetic energy of bodily moments.) The conservation of energy, however, implied that such fundamental special forces must be governed by strict deterministic laws: if mental or vital forces arose spontaneously, then there would be nothing to ensure that they never led to energy increases. This had a great impact on 19th-century thought about free will. The idea that all mental and vital processes must be entirely governed by deterministic laws was viewed by many as incompatible with the traditional view of free agents as autonomous influences operating independently of the constraints of natural law.

It is an interesting question whether the 19th-century view that all fundamental mental or vital forces must be governed by deterministic laws already amounts to a doctrine worth calling "physicalism." In favor of this way of viewing things, the doctrine does portray all causally significant properties, including mental and vital ones, as within the realm of scientific theory and in principle subject to the kind of mathematical treatment familiar from the analysis of gravity and electromagnetism. But on the other side, the mental and

4 For discussion of the compatibility of fundamental mental forces with the conservation of energy law, see Montero (2006).

vital forces that it allows are fundamental causal agents, found only in sentient and living organisms, and additional to any forces operating in the inanimate world. Some will feel a doctrine that countenances such fundamental mental and vital forces is not worth counting as "physicalist."

Strong Physicalism Vindicated

As it happens, there is no great need to determine whether being governed by mathematically formulable deterministic laws suffices for an entity to count as physical. This is because 20th-century science has given us reason to suppose that there are no fundamental vital or mental forces after all, and that the only things capable of producing physical effects are entities constituted by the kind of basic force fields that can be found throughout the inanimate world (such as gravity, electromagnetism, and nuclear forces). Such science thus supports the stronger view that everything is made of entities that are not only governed by deterministic mathematical laws, but are also all found in inanimate realms. In other words, it supports the view that the underlying nature of humans and other creatures with minds in not different in kind from the underlying nature of inanimate things such as rocks. (In line with this, and given that the term "physical," when used in debates over physicalism, is a philosophical term of art, we will now stipulate that physical entities are those found in inanimate realms and those composed out of such entities.)

The argument against fundamental vital and mental forces is a simple empirical one. By the 1950s, it had become difficult to continue to uphold the existence of special vital or mental forces: detailed physiological research, especially into nerve cells, gave no indication of any physical effects that cannot be explained in terms of the basic physical forces that also occur outside living bodies. A great deal became known about biochemical and neurophysiological processes, especially at the level of the cell, and none of it gave any evidence for the existence of special forces not found elsewhere in nature.

Thus, during the first half of the century, the catalytic role and protein constitution of enzymes were recognized, basic biochemical cycles were identified, and the structure of proteins was analyzed, culminating in the discovery of DNA. In the same period, neurophysiological research mapped the body's neuronal network and analyzed the electrochemical mechanisms responsible for neuronal activity. Together, these developments made it difficult to go on maintaining that special forces operate inside living bodies. If there were such forces, they could be expected to display some manifestation of their presence. But detailed physiological investigation failed to uncover evidence of any nonphysical forces. The underlying nature of human beings, this research indicated, is no different from the underlying nature of ordinary nonliving things (Papineau 2001).

This argument rests on normal inductive grounds. There is no principled *a priori* reason why 20th-century physiological research should not have uncovered special mental and vital forces. It is just that the inductive evidence went the other way. Of course, it is possible to resist the conclusion. You could continue to believe that there are special vital mental forces that operate in as yet undetected ways in the interstices of living tissues and intelligent brains, and resist physicalism on those grounds. But there seems little merit in this position. The nonexistence of special vital and mental forces has been established by over a century of detailed empirical research. Given this, it seems more rational to explore the consequences of this finding, rather than resist it.

Hempel's Dilemma

Some philosophers seek to cast doubt on physicalism by appealing to "Hempel's dilemma." In the 1950s, Carl Hempel argued that physicalism falls at the first hurdle on the grounds that there is no good way of understanding the term "physical." We might take "physical" to refer to just those entities that are recognized by contemporary physics. But then the doctrine that *everything is physical* will certainly be false, for we can be confident that contemporary physics is by no means the last word about the basic constituents of reality. Alternatively, we might understand "physical" as referring to just those entities recognized in the future by the successors of our contemporary theories, at the ideal end of enquiry, perhaps. But then the doctrine that *everything is physical* will be empty, for we have little idea what future physics will reveal.[5]

However, notwithstanding Hempel's suggestion, there is no need to define "physical" in terms of physical theory, either contemporary or future, to pick out some privileged notion of "physical." In truth, there are a number of *alternative* ways of defining "physical", *all* of which give rise to interesting doctrines arguably worth calling "physicalism." All that is needed is *some* way of identifying a category of facts (call them "Q") that satisfies the following requirements: (1) at first sight mental, biological, and similar categories do not seem to be Q, but (2) mental, biological, and similar facts nevertheless do have effects among Q-facts, while at the same time (3) there is good reason to think that Q is "causally complete": that is, that Q-effects always have fully Q-causes. As soon as we have a Q-category that satisfies these specifications, we can illuminatingly argue as before that mental, biological, and similar categories must after all be composed of Q-facts, despite first appearances – for otherwise how could they cause their Q-effects, given the completeness thesis that Q-effects always have fully Q-causes?[6]

As it happens, we have already identified *two* different categories satisfying these specifications. We explained earlier how the 19th-century discovery of the conservation of energy showed that the category of facts "falling under deterministic mathematically formulable laws" satisfies a completeness requirement, in that the discovery implied that all effects of this kind must have similar causes (in the form of conservative force fields). The implication was thus that mental and other such facts must themselves fall under deterministic mathematical laws, despite initial appearances to the contrary. And then we pointed out that 20th-century physiology gives us strong reason to think that the realm of phenomena composed of the kinds of entities found in inanimate realms is also causally complete, in that even within brains and bodies effects always seem to be produced by just the same kinds of electrochemical and other causes as operate in the inanimate realm. And this then argued that mental and other features of the world that have inanimate effects, in addition to falling under deterministic mathematical laws, must also be composed of the kinds of entities found in the inanimate realm.

Other philosophers run this form of argument with yet other specifications of a Q-category, such as "determined by microscopic components" or, again, "similar to the kind of entities recognized by current physics." To the extent that these categories satisfy the relevant specifications, and in particular a completeness requirement that effects of these kinds always have full causes of these kinds, they allow further interesting conclusions.

5 For discussion of Hempel's dilemma, see Montero (1999).

6 See Papineau and Spurrett (1999).

Do all of these conclusions line up with views that have traditionally been thought of as forms of physicalism? Some do, some might not, but, as we said earlier, "physicalism" is a term of philosophical art that philosophers use in a variety of ways. Our point is merely that any category that fits our criteria can be plugged into the causal argument to produce surprising results. Hempel's dilemma is avoided because the role of science, including physical science, is not, on our account, to give a definition of "physical," but rather to tell us whether the so-defined "physical" realm is causally complete. And it turns out that science supports the completeness claim for a range of differently defined "physical" realms.

It is important to realize that science does not have to tell us everything in order to tell us anything. Science has not yet, of course, verified a complete definitive list of the fundamental entities responsible for effects in the natural world. But this does not mean that it has not yet verified any significant facts about that list. And our historical analysis indicates that it has indeed established a significant amount of such information; for example, that those fundamental agents are all governed by deterministic mathematical laws and all operate inter alia in the inanimate realm (and plausibly also that they are all determined by microscopic components and all similar to the kinds of entities recognized by current physics).

So we have a number of alternative possible "physicalisms" to choose among, all of which give us highly interesting conclusions about the constitution of the world, and in particular about the constitution of mental and biological agents. In the interests of clarity and brevity, however, let us just stick to the definition we settled on earlier, and continue to understand "physical" as referring specifically to items composed of entities that are also found within the inanimate world.

The Rise of Physicalism in the 1950s and 60s

It is striking how physicalism emerged suddenly as a popular philosophical doctrine in the middle decades of the 20th century. From the 1950s on, a number of philosophers put forward arguments designed to establish physicalist conclusions, and in particular to show that the mind must be identified with the brain. Not all these arguments may seem to fit the analysis of physicalism we have given so far. While some of them did appeal explicitly to the causal closure of the physical realm (Feigl 1958; Oppenheim and Putnam 1958), other arguments from this period seemed to proceed rather differently.

However, it is not difficult to show that appearances are deceptive here. On examination, it turns out that even those arguments that don't explicitly appeal to causal closure do so implicitly, in line with our thesis that it was the empirical evidence for causal closure that persuaded philosophers to be physicalists. Once mid-century physiological research had established that all physical effects had physical causes, even in bodies and brains, philosophers quickly figured out that general physicalism followed (Papineau 2001).

Thus, for example, consider Smart's (1959) thought that we should identify mental states with brain states, for otherwise those mental states would be "nomological danglers" that play no role in the explanation of behavior. Similarly, reflect on Lewis (1966) and Armstrong's (1968) arguments that, since mental states are picked out by their causal roles, including their roles as causes of behavior, and since we know that physical states play these roles, mental states must be identical with those physical states. Or, again, consider Davidson's (1970)

argument that, since the only laws governing behavior are those connecting behavior with physical antecedents, mental events can only be causes of behavior if they are identical with those physical antecedents. There is much to say about these arguments. But the point we want to make here is that none of them is even slightly plausible without the assumption of the causal closure of physics.

To see this, imagine that the causal closure of physics were not true, and that some physical effects (the movements of arms, perhaps, or the firings of the motor neurons that instigate those movements) were not determined by law by prior physical causes at all, but instead by fundamental dualist mental causes, such as decisions or exercises of will, or perhaps just pains. If this were the case, then (1) contra Smart, mental states wouldn't be "nomological danglers," but would be directly efficacious in the production of behavior; (2) contra Lewis and Armstrong, it wouldn't necessarily be physical states that played the causal roles by which we pick out mental states, but quite possibly the *sui generis* mental states themselves; and (3) contra Davidson, it wouldn't be true that the only laws governing behavior are those connecting behavior with physical antecedents, since there would also be laws connecting behavior with mental antecedents.

Arguments against Physicalism

The contemporary philosophical literature contains a number of arguments against physicalism about the mental realm, and in particular about *conscious* mental states. Should they make us reconsider our stance on physicalism?

Central to the antiphysicalist literature is Frank Jackson's "knowledge argument" (1982), which takes up themes from Thomas Nagel's influential earlier article "What Is it Like to Be a Bat?" (1974). These authors point out that someone could know all there is to know from a third-person scientific point of view about certain conscious states yet not "know what it is like" to undergo them. Thus, Nagel observed that even an expert bat scientist could not know what it is like to echolocate, and Jackson constructed a thought experiment in which a fully scientifically informed but color experience-deprived vision scientist could not know what it is like to see colors.

Whether these examples discredit physicalism hinges on what is involved in "knowing what it is like" to undergo some experience. If this required knowing about some nonphysical feature of the world, then physicalism would certainly be false.

But it is contentious whether talk of "knowing what it is like" carries this implication. Nearly all contemporary physicalists allow that someone who has never undergone a certain kind of experience is cognitively limited as a result, and that a color experience-deprived vision scientist would learn something knew when she saw color for the first time. But they do not concede that this change involves her coming to know about something nonphysical. There are many subtle differences in how physicalists try to explain the change in the scientist. However, they all agree in aiming to do so without bringing in any nonphysical properties, instead appealing solely to the entirely physical, neural changes that ground the experience of seeing color.[7]

At this point, many antiphysicalists introduce a further line of argument. They insist that, if you "know what some experience is like,", then your introspective thinking about

7 For two such accounts, see Papineau (2002) and Montero (2007).

that experience will be manifestly "transparent," in that it will be guaranteed to reveal the essential properties of the experience. (When you think about an experience in terms of "what it is like," so the thought goes, isn't its real nature made apparent to you?) But such introspective thought does not show experiences to be physical states, point out the antiphysicalists, who conclude that they are not physical.

In response, physicalists will deny that the introspective mode of thinking about experiences is transparent in the relevant sense. While they maintain that experiences do have a physical nature, they say that there is no reason why introspective thinking should reveal this immediately. If physicalism is true, the mental is entirely grounded in inanimate aspects of the world. When we introspect, on the physicalist's view, we are aware of the mental but not aware of what grounds it.

Still, many feel that there is something very counterintuitive about a physicalist view of consciousness, and no doubt this lends plausibility to the antiphysicalist arguments. But an intuition is not an argument, and in any case physicalists have offered a number of suggestions to account for the intuitive difficulty of embracing physicalism (see Papineau 2011).

In the end, the strongest reason for distrusting an antiphysicalist view of consciousness is the strength of the contrary argument we have already examined in favor of physicalism. Somebody who denies physicalism about consciousness needs to find some fault in this argument, and the options are limited. We shall now examine these options.

The Causal Argument Analyzed

The defense of physicalism we have considered so far has come to be known in the philosophical literature as "the causal argument." To repeat, the general idea of the argument is this: a nonphysical mind would be like a ghost in a machine that has the power to flip switches and thereby cause our physical bodies to move. However, we have good reason to believe that all of these machine switches are flipped on or off by other physical parts of the machine. And since it is absurd to think that the switches are doubly flipped by both the machine and the ghost, we should conclude that there is no ghost in the machine, that the mental causes of our bodily movements are themselves physical parts of the machine. To help us explore the alternatives open for those who wish to resist physicalism, let us lay out the argument formally with three premises:

(1) Mental (and biological, etc.) states have physical effects.
(2) All physical effects have full physical causes.
(3) The physical effects of conscious causes aren't systematically overdetermined by two or more distinct causes.

Given these premises, the conclusion that the mental (biological, etc.) states mentioned in premise (1) must be part of the physical causes mentioned in premise (2) is inescapable. (After all, (1) tells us certain effects have mental (biological, etc.) causes, (2) tells us those effects have full physical causes, and (3) tells us those effects don't systematically have two or more distinct causes. The only way for these all to be true is for the mental (biological, etc.) causes to already be included in the physical causes.)

So someone who wants to deny physicalism about conscious mental states needs to deny one of the premises. What are the options?

Epiphenomenalism and Preestablished Harmony

A first option is to deny (1). You can hold that, despite first appearances, conscious mental states like pains, feelings, and decisions do not in fact cause bodily movements or any other effects in the physical world.

This position is most familiar in the guise of *epiphenomenalism*. Epiphenomenalists view the conscious mind as an inefficacious side-effect of the brain's operations. They agree with physicalists that bodily movements and so on are fully accounted for by brain processes, but insist that the conscious mind floats above these brain processes, as it were, rather like the puffs of smoke that are emitted by a steam train, but which themselves make no causal contribution to the train's progress.

An alternative way of denying (1) is to adopt Leibniz's doctrine of *preestablished harmony*. On this view, not only does the mind not affect the body, but the body doesn't affect the mind either. Rather, God has set both to run along parallel tracks, marching in perfect step so that we remain unaware of their causal independence.

The basic objection to both these doctrines is that they are theoretically quite implausible. Epihenomenalism requires us to suppose that conscious states, even though they are caused by processes in the physical world, have no effects on that world. This is a very odd kind of causal structure. Nature displays no other examples of such one-way causal intercourse between realms. Preestablished harmony is, if anything, even odder. Again, there are no other examples of nature dividing itself into causally isolated but coordinated realms. Normal principles of theory choice would seem to argue strongly against such convoluted causal structures, and in favor of the far simpler physicalist position that integrates the mental realm with the causal unfolding of the spatiotemporal world in an entirely familiar way.

If we are going to appeal to a principle of simplicity at this point, we might start wondering what work is being done by the causal argument laid out in the previous section, and in particular by its second causal closure premise. If general principles of theory choice can justify physicalism, why bring in all the complications associated with causal closure? The answer is that causal closure is needed to rule out interactionist dualism. General principles of theory choice may dismiss epiphenomenalism in favor of physicalism, but they do not similarly discredit interactionist dualism. As the brief historical sketch earlier will have made clear, interactionist dualism offers a perfectly straightforward theoretical option, requiring no commitment to any bizarre causal structures. Certainly, the historical norm has been to regard it as the default account of the causal role of the mental realm. Given this, arguments from theoretical simplicity are ineffective against interactionist dualism. Rather, the case against interactionist dualism hinges crucially on the empirical thesis that all physical effects have full physical causes. It is specifically this claim that makes it difficult to see how dualist states can make a causal difference to the physical world.

It is sometimes suggested that physicalism about the mind can be vindicated by an "inference to the best explanation." The thought here is that there are many well-established synchronic correlations between mental states and brain states, and that physicalism is a "better explanation" of these correlations than is epiphenomenalism (Hill 1991; Hill and McLaughlin 1999; Melnyk 2003). From the perspective outlined here, this starts the argument in the middle rather than the beginning, by simply assuming the relevant mind–brain correlations. This assumption of pervasive synchronic mind–brain correlations is only plausible if interactionist dualism has already been ruled out. After all, if we believed

interactionist dualism, then we would have no particular reason to think that every mental state is systematically correlated with some particular kind of brain state – since mental states, on the interactionist view, do not depend on any underlying neural processes.

Denying Causal Closure

These last points immediately indicate another possible response to the causal argument for physicalism. You could deny the second causal closure premise and embrace interactionist dualism. As observed earlier, there is nothing theoretically unattractive about this option. As far as initial plausibility goes, many see interactionism as preferable to physicalism. The case against it rests not on the superior plausibility of physicalism, but on the empirical case for causal closure.

It is noteworthy that almost no contemporary theorists defend interactive dualism. There were still a number of serious thinkers 50 years ago – the philosopher Karl Popper and the neuroscientist John Eccles spring to mind – who endorsed the view that nonphysical factors operated in the brain to influence its physical operations. But nowadays such views have few advocates in scientific or philosophical circles. We take this to testify to the strength of empirical evidence that built up through the 20th century against the existence of any nonphysical force fields. (Figures like Popper and Eccles were not so much eccentric as *old*: they grew into intellectual maturity at a time when mental and vital forces were widely taken for granted.)

It might occur to some readers that, if we are focusing on 20th-century science, then the indeterminism of modern quantum mechanics surely counts against the causal closure of the physical realm, and therewith undermines the causal argument for physicalism. Doesn't quantum mechanics show us that plenty of physical effects are chancy, and so don't have full physical causes? And doesn't this then leave room for an independent nonphysical mind to come in and affect what happens in the physical world?

This objection, however, is readily addressed. Even if quantum mechanics implies that some physical effects are themselves undetermined, it provides no reason to doubt a quantum version of the causal closure thesis, to the effect that the *chances* of those effects are fully fixed by prior physical circumstances. And this alone is enough to rule out any role for nonphysical causes. Such nonphysical causes, if they are to be genuinely efficacious, must make an independent difference to the chances of physical effects, and this in itself would be inconsistent with the quantum causal closure claim that such chances are already fixed by prior physical circumstances. Once more, it seems that anything that makes a difference to the physical realm must itself be physical.

It is striking that even those contemporary philosophers who are persuaded by the antiphysicalist arguments about consciousness do not typically respond by denying causal closure and embracing interactionism. Rather, they tend to go for epiphenomenalism.[8] Both Frank Jackson and David Chalmers initially combined their rejection of physicalism

8 But see Lowe (2000; 2003), who suggests that there may be "invisible" violations of causal closure, due to the irreducible mental causes of my bodily behavior always themselves having physical causes. On Lowe's view, while those bodily behaviors might *seem* to follow by physical law from the physical antecedents of their mental causes, they would not in fact ensue in possible worlds that shared all our physical laws but lacked the extra mental processes that occur in this world. Our response to such a position is, again, that science has failed to turn up any evidence for physical behaviors that cannot be fully accounted for in terms of prior physical processes alone.

with an advocacy of epihenomenalism (though it should be said that the difficulties of epiphenomenalism have since persuaded them to move away: Jackson is now a physicalist, while Chalmers favors a "neutral monism" whose differences from physicalism are a matter of debate).[9]

Denying Systematic Overdetermination

One final way for nonphysicalists to evade the causal argument would be to deny premise (3), and so have the physical effects of conscious mental causes systematically caused twice over, both by a brain process and by an independently efficacious mental state. However, while this "belt and braces" option has had one or two defenders (Crane and Mellor 1990; Mellor 1995), it seems just as open to the accusation of unnecessary complexity as epiphenomenalism, if not more so. While occasional overdetermination by independent causes does sometimes occur (a man might be shot and struck by lightning simultaneously, for example), nature does not seem to offer any other examples of a whole *category* of effects that is *systematically* overdetermined by two independent causes. (Not to mention that the "belt and braces" view would seem to be in need of some additional mechanism to ensure that both causes are in place whenever one of them is.)

It is worth distinguishing this kind of "vicious" overdetermination from a species of "benign" overdetermination favored by some advocates of nonreductive physicalism. In order to explain this point, it will be necessary to backtrack a little.

Recall our initial characterization of physicalism. We were at pains to explain that it is a relatively permissive doctrine, and is not committed to the thesis that all properties, including mental properties, can be defined in the language of basic physics. Rather, the essential core was only the "metaphysical supervenience" of all the facts on the basic physical facts: once the basic physical facts are fixed, nothing more is needed for all the facts to be in place.

Let us use the term "*reductive* physicalism" for the stronger doctrine that all properties, including mental properties, can be identified with properties that are definable in the language of basic physics. Then "*non*reductive physicalism" is the weaker claim that all the facts supervene on the basic physical facts or are basic physical facts themselves. Such a view allows for "multiple realization"; that is, it allows for higher-level properties, including mental properties, to be composed by different lower-level physical properties in different cases. Pain, on this view, is not identical to a certain type of neural process, such as C-fiber stimulation, but can be determined by, or supervene on, different types of physical process in different kinds of creature. Such a view allows for extraterrestrials, if there are any, to experience pain while having very different underlying physical natures than we do.

Most contemporary physicalist philosophers probably favor nonreductive over reductive physicalism. But this, as some see it, generates a problem. If pain, for example, is not strictly identical to some lower-level physical cause such as C-fiber stimulation, then every time your C-fibers cause you to scream out and writhe, your pain would seem to act as another cause of your behavior, and thus it appears that the physical effects of mental causes are, after all, doubly caused. Thus, nonreductive physicalism might seem to saddle us with unacceptable proliferation of overdetermining causes for the physical effects of mental causes after all: both the physical cause implied by the causal closure thesis and the distinct mental cause.

9 See Montero (2015).

Advocates of nonreductive physicalism have a response: as they see it, there is nothing wrong with such an apparent duplication of causes if it is also specified that one cause metaphysically supervenes on the other. The issue here hinges on the acceptability of different kinds of overdetermination (Bennett 2003; Melnyk 2003). All can agree that it would be absurd if the physical effects of nonphysical causes always had two completely independent causes. However, even if such "vicious overdetermination" by two ontologically independent causes is so ruled out, as assumed by the causal argument, this does not necessarily preclude "benign overdetermination" by both a physical cause and a metaphysically supervenient mental cause, argue advocates of nonreductive physicalism. In their view, this kind of overdetermination is acceptable, on the grounds that the two causes are not ontologically distinct: the nonphysical cause isn't additional to the physical cause (nothing more is needed for your feelings than your brain states).

Not everybody agrees that this solution is satisfactory. A substantial minority among physicalists feel that the causal status of nonreduced mental categories is suspect, and so conclude that the only good physicalism is reductive physicalism. This is not the place to adjudicate this technical dispute.

The important point for present purposes is to note that this is an internal dispute among physicalists, rather than an argument against physicalism. It is agreed on all sides that the causal closure of the physical requires *at least* the metaphysical supervenience of the mental on the physical, on pain of denying any causal efficacy to mental states or positing a systematic vicious overdetermination. The only further question is whether it also requires the strong reduction of mental properties to physical ones, or whether it is acceptable to leave the mental properties as merely metaphysically supervenient.

As we said at the beginning, "naturalism" means different things to different people. But one central question raised by the term is whether mental, biological, and other such entities are all physical. We have tried to show how this strand in naturalism has strong scientific support.

References

Armstrong, D. (1968). *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.

Bennett, K. (2003). Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract It. *Noûs* 37: 471–497.

Crane, T. and Mellor, D. (1990). There is No Question of Physicalism. *Mind* 99: 185–206.

Davidson, D. (1970). Mental Events. In *Experience and Theory*, edited by L. Foster and J. Swanson. London: Duckworth.

Elkana, Y. (1974). *The Discovery of the Conservation of Energy*. London: Hutchinson.

Feigl, H. (1958). The "Mental" and the "Physical." In *Minnesota Studies in the Philosophy of Science*, vol. II, edited by H. Feigl, M. Scriven, and G. Maxwell. Minneapolis: University of Minnesota Press.

Hill, C. (1991). *Sensations*. Cambridge: Cambridge University Press.

Hill, C. and McLaughlin, B. (1999). There are Fewer Things in Reality than are Dreamt of in Chalmers' Philosophy. *Philosophy and Phenomenological Research* 59: 445–454.

Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly* 32: 127–136.

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

Lewis, D. (1966). An Argument for the Identity Theory. *Journal of Philosophy* 63: 17–25.

Lowe, E.J. (2000). Causal Closure Principles and Emergentism. *Philosophy* 75: 571–585.

Lowe, E.J. (2003). Physical Causal Closure and the Invisibility of Mental Causation. In *Physicalism and Mental Causation*, edited by S. Walter and H.-D. Heckmann. Exeter: Imprint Academic.

Mellor, D. (1995). *The Facts of Causation*. London: Routledge.

Melnyk, A. (2003). *A Physicalist Manifesto*. Cambridge: Cambridge University Press.

Montero, B. (1999). The Body Problem. *Noûs* 33: 183–220.

Montero, B. (2006). What Does the Conservation of Energy Have to Do with Physicalism? *Dialectica* 60: 383–396.

Monetero, B. (2007). Physicalism Could Be True Even if Mary Learns Something New. *Philosophical Quarterly* 57: 176–189.

Montero, B. (2013). Must Physicalism Imply the Supervenience of the Mental on the Physical? *Journal of Philosophy* 110: 93–110.

Montero, B. (2015). Russellian Physicalism. In *Consciousness in the Physical World: Perspectives on Russellian Monism*, edited by Torin Alter and Yujin Nagasawa. Oxford: Oxford University Press.

Nagel, T. (1974). What Is it Like to Be a Bat? *Philosophical Review* 82: 435–450.

Oppenheim, P. and Putnam, H. (1958). Unity of Science as a Working Hypothesis. In *Minnesota Studies in the Philosophy of Science*, vol. II, edited by H. Feigl, M. Scriven, and G. Maxwell. Minneapolis: University of Minnesota Press.

Papineau, D. (2001). The Rise of Physicalism. In *Physicalism and Its Discontents*, edited by B. Loewer and C. Gillett. Cambridge: Cambridge University Press.

Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.

Papineau, D. (2011). What Exactly is the Explanatory Gap? *Philosophia* 39: 5–19.

Papineau, D. and Spurrett, D. (1999). A Note on the Completeness of "Physics." *Analysis* 59: 25–29.

Smart, J. (1959). Sensations and Brain Processes. *Philosophical Review* 68: 141–156.

Woolhouse, R. (1985). Leibniz's Reaction to Cartesian Interaction. *Proceedings of the Aristotelian Society* 86: 69–82.