able to see how pain causes distress, since we have from neurophysiology a detailed story about how N₁ causes N₂. I am suggesting this as a way of understanding what the inference (9) above shows. Viewed this way, (9) is a straightforward causal argument for reductionism, or type physicalism: if mental kinds are identical with physical kinds, that will vindicate mental causation.

But it is important to see that this is not at all a conclusive argument for type physicalism—for two reasons. First, even if we assume that mental causation is real, we would have to show that type identities are the only way of showing mental causation to be possible. It may be true—I think it is true—that type physicalism will vindicate mental causation, but it may not be the only position on the mind-body problem that can do this. In my view, functional reduction of the sort I have discussed, which, unlike Block and Stalnaker’s type physicalism, is immune to the notorious multiple realization argument, can also ground mental causation. Second, even if we are entitled to the conditional “Mental causation is possible only if some form of reductionism holds,” we cannot infer that reductionism is true. For we should not a priori rule out epiphenomenalism and other noncausal views of the mind. Reductionism must be earned—and mental causation, too, must be earned—by showing that mental properties are indeed reducible, in a relevant sense, to physical/biological properties. And this requires independent arguments and evidence. Mental causation is a presumptive desideratum, and it may be very high on our wish list. But our wish to save mental causation, however sincere and righteous, cannot by itself make reductionism true.
aggregated out of bits of matter, all behaving in accordance with laws of physics, and that any phenomenon of the world can be physically explained if it can be explained at all. When physicalism is accepted as a basic framework, the foremost metaphysical question about the mind is where in the physical world our minds and mentality fit—indeed, whether minds have a place in a physical world at all. In this chapter, I want to discuss just where we are with the physicalist program, in order to see to what extent physicalism’s hopes and promises have been, or are likely to be, realized or frustrated—that is, how much, or what kind of, physicalism we can have. As we will see, the efforts do not succeed entirely, but I will claim that they succeed nearly enough. But first we need to review just where we are with the claims and arguments so far advanced in this book.

**Taking Stock**

The most fundamental tenet of physicalism concerns the ontology of the world. It claims that the **content of the world is wholly exhausted by matter**. Material things are all the things that there are; there is nothing inside the spacetime world that isn’t material, and of course there is nothing outside it either. The spacetime world is the whole world, and material things, bits of matter and complex structures made up of bits of matter, are its only inhabitants. This doctrine is sometimes called “ontological physicalism.”

But why should we accept ontological physicalism? Why can’t there be things other than material things? In an earlier chapter a causal argument was mounted against Cartesian mental substances, minds conceived as concrete immaterial things outside physical space in causal interaction with material things located within physical space. We saw that if minds were such mental substances, they could not possibly causally influence, or be influenced by, material things, and further, they could not causally interact with one another either. That is to say, immaterial nonspatial minds would be totally causally impotent, and this renders them explanatorily irrelevant and useless. Moreover, such a radically noncausal view of minds makes it difficult to understand how we could ever come to know that there are minds. Our considerations do not show that causal relations cannot hold within a single mental substance (even Leibniz, famous for disallowing causation between monads, allowed it within a single monad, or so I understand). However, what has been shown, I believe, is sufficient to defeat any rationale for substance dualism. Causality requires a domain with a space-like structure—that is, a “space” within which objects and events can be identified by their “locations”—and, as far as we know, the domain of physical objects is the only domain with a structure of that kind. If this is right, we have a causal argument for monistic physicalist ontology.

Although substance dualism has not been taken as a serious option in mainstream philosophy of mind since the early twentieth century, a general dualist perspective has shown itself to possess a surprisingly tenacious capacity for survival. The dualism of mental and physical properties, in various forms, has dominated the field during the second half of the last century, and continues to hold the allegiance of a majority of philosophers, although it is fair to say that what Ned Block has aptly called “antireductionist consensus,” which was firmly in place by the mid-1970s, is no longer what it used to be. I believe the persistence of dualism points to what strikes me as a deeply entrenched, almost instinctive, aversion that most of us feel toward reductionist physicalism. Most of us would like to believe, it seems, that although we are wholly composed of bits of matter, we are extraordinarily complex physical systems with properties, capacities, and functions that are not merely physical or reducible to the merely physical. We are apt to feel that reductive physicalism, according to which thoughts and
feelings are mere molecular movements, cannot do justice to the special and distinctive position that we occupy in the natural order, and that a reductionist physicalist is in self-denial, a denial of his own distinctive and unique nature. An idea like this is found not only in those wedded to a traditional, often religiously inspired, conception of ourselves as persons; it also seems to be the driving force behind the widely shared view that the special sciences, especially those concerning humans, like psychology, cognitive science, and economics, are autonomous and irreducible to the physical and biological sciences. Current antireductionism as it concerns the special sciences seems almost like a replay of the influential doctrine in late nineteenth-century Europe that posited a fundamental difference between the Naturwissenschaften and the Geisteswissenschaften—that is, between the natural sciences and the humane sciences. In any case, the long reign of nonreductive physicalism is a testimony to its appeal as a thesis about the nature of the special sciences as well as a position on the mind–body problem. It promises to safeguard both the autonomy of the special sciences and our specialness as cognizers and agents. But philosophical positions cannot live on hopes and messages alone: they have to deliver the goods as advertised.

As it has turned out, nonreductive materialism could not deliver on mental causation—any better than Cartesian dualism could. It could not explain, on its own terms, how mental phenomena, like belief, desire, feeling, and sensation, could causally affect the course of events in the physical world. Mind–to–body causation is fundamental if our mentality is to make a difference to what goes on in the world. If I want to have the slightest causal influence on anything outside me—to change a light bulb or start a war—I must first move my

limbs or other parts of my body; somehow, my beliefs and desires must cause the muscles in my arms and legs to contract, or cause my vocal cords to vibrate. Mental causation is fundamental to our conception of mentality, and to our view of ourselves as agents and cognizers; any theory of mind that is not able to accommodate mental causation must be considered inadequate, or at best incomplete.

So how can a mental phenomenon, say, my desire for a drink of water, manage to cause my legs to move so as to transport my entire body to the kitchen? Earlier we have discussed in considerable detail an argument, the “supervenience” or “exclusion” argument, whose conclusion is that causally efficacious mental phenomena must be reducible to physical ones, and, more broadly, that, given the closed character of the physical domain, any phenomenon that is causally linked with a physical phenomenon must itself be a physical phenomenon. If mental phenomena are neural processes in the brain, there will be no special mystery about mental causation; I believe we already know the neurophysiology involved well enough—how neural excitations in the motor cortex send electrochemical signals down through the efferent nerve channels to the appropriate muscles, causing them to contract, which in turn causes the limbs to move. According to property dualism, however, mental phenomena are distinct from neural phenomena, and it becomes a prima facie mystery by what mechanisms, or through what intervening links, these supposedly nonphysical phenomena can cause the muscles to contract. At least, this much seems undeniable: If my desire is to cause my legs to move, it must somehow make use of, or ride piggyback on, the causal chain from my motor cortex to the leg muscles. It just is not possible to believe that my desire might be able to act on my legs directly, through some form of telekinesis, or that there could be another causal path, independent of the neural/physical causal chain, that connects the desire to the leg movement. It seems then that if a neural

event in the motor cortex is the ultimate physical cause of the leg movement, my desire must somehow cause that neural event. But how is that possible?

Much is known about the physicochemical processes involved in the firing of a bundle of neural fibers—how electrical potential builds up in a neuron until it reaches a critical point and then discharges. The rising of electric potential involves the movement of electrically charged molecules; if some nonphysical causal agent is to cause a neuron to fire, it must be able to causally influence the motion of molecules. I assume we know something about how molecular motions take place inside a cell, what physical forces are active in causing these motions, and just how the motions depend on the magnitudes of these forces. Do we really have any conception of how some immaterial, nonphysical force might change the motion of even a single molecule, causing it to speed up or slow down, or change its direction? Will brain scientists ever look for nonphysical forces, or nonphysical phenomena, to explain some neural event for which they are having difficulty identifying a physical cause? If they should decide to do that, how would they go about it? Where and how would they look for nonphysical causes of neural events? How would they identify one and measure its properties?

The answer of course is that brain scientists will not look outside the physical domain for explanations of neural phenomena. They are not likely to think that it will be scientifically productive to look for nonphysical, immaterial forces to explain neural events. We expect the physical world to be causally self-contained and explanatorily self-sufficient. That is, we suppose that if a neural event—or more broadly, a physical event—has a cause, or an explanation, then it must have a physical cause and a physical explanation. This is the principle of causal/explanatory closure of the physical domain.

Some dualists may think that these considerations are question-begging in that they assume the causal closure of the physical. There are excellent, even compelling, reasons for accepting the causal and explanatory self-sufficiency of the physical world, but rather than arguing this point, let me show you another way of generating the problem of mental causation for property dualists. Consider a mental event, say an occurrence of pain. We believe—in fact, we know—that pain occurs only because a certain neural state, call it $\psi$, occurs. $\psi$ may differ from organism to organism, especially in organisms belonging to different biological species, but we do not think that there are sensations that float free from the brain, without being grounded in underlying neural processes. Assume that $\psi$ is the neural substrate of pain in you. If $\psi$ occurs, you will experience pain, and you will not experience pain unless $\psi$ occurs. Consider the claim that the pain caused your finger to twitch. Suppose, further, that neurophysiologists have discovered a causal chain from $\psi$, the neural substrate of your pain, to the finger twitching, establishing $\psi$ as its sufficient physical cause. I believe the existence of such a causal chain is highly likely; we already know a lot about physical causal processes underlying many sensory processes. This means that your finger twitching has two putative causes, one mental (your pain) and one physical (the pain's neural substrate $\psi$). Given that your finger twitching, a physical event, has a full physical cause, how is a mental cause also possible? How could one and the same event have two distinct causal origins? Doesn't the physical cause threaten to preempt the supposed mental cause? This is the problem of causal exclusion, as may be recalled. Note that in developing this problem, we have used various highly plausible empirical assumptions about the neurophysiology of pain (along with some commonsensical assumptions about causality), but not the principle of physical causal closure.

It is clear that property dualism has no way of dealing with these questions, and that, especially in view of physical causal closure, phenomena outside the physical domain must remain causally impotent, mere epiphenomena, at least with regard to physical phenomena. In 1970, Donald Davidson's "Mental Events" appeared, and this influential paper sparked the revival of the problem of mental causation, more than three hundred years after the problem doomed Descartes's interactionist dualism. What has become increasingly evident over the past thirty years is that mental causation poses insuperable difficulties for all forms of mind-body dualism—for property dualism no less than substance dualism. Some philosophers are still gamely holding on, trying to somehow fashion an account of mental causation within the nonreductive scheme, but I believe that if we have learned anything from the three decades of debate, it is the simple point that unless we bring the supposed mental causes fully into the physical world, there is no hope of vindicating their status as causes, and that the reality of mental causation requires reduction of mentality to physical processes, or of minds to brains.

**Physicalism at a Crossroads**

In getting to where we are at this point, the first choice point we faced was where we had to decide between substance dualism, which posits both material bodies and immaterial minds, and ontological physicalism, which admits only material objects. Motivations for introducing entities other than material things vary—from supposed philosophical requirements in connection with certain issues, for example, the persistence of persons over time, a possible survival of bodily death, and the special directness of knowledge of one's own mind, to religious imperatives and mystical intimations. As we saw, however, immaterial minds, outside physical space, are causally cut off not only from physical objects and events but also from other minds as well. Each immaterial mind would be a totally isolated entity; its existence is inexplicable and its presence or absence can make no difference to anything else. I believe this is sufficient ground for rejecting substance dualism of all kinds, or any ontology that accedes to immaterial, nonphysical things. The alternative is physicalism with an ontology countenancing only material things. There is just one kind of substance, and it is material substance. The physical world of spacetime is the whole world.

Let us suppose then that at the first choice point we have opted for ontological physicalism. The next choice point we face concerns the properties of material things: Given that only material things exist, what kinds of properties do they, or can they, have? Specifically, can they have properties that are not physical? That is, can they have properties that are not dealt with in fundamental physics or reducible, in some broad but clear sense, to fundamental physical properties? As we saw, many thinkers have been attracted to an affirmative answer to this question, holding the view that complex physical systems can exhibit novel, "emergent" properties that are not reducible to the properties of their simpler constituents. Early emergentists thought that even such simple physicochemical properties as the transparency of water were emergent; now it is generally conceded that consciousness and mentality are the best, and perhaps the only, candidates for genuine emergence. Emergentism, which is now showing signs of a revival after having been moribund for much of the second half of the twentieth century, is a form of property dualism, the position that, in

---


4. What of abstract things, like numbers, properties, propositions, and the like? I am concerned here only with the concrete world; I am setting aside the issue of Platonic objects, although some philosophers believe that Platonism is excluded by naturalism.

5. There have recently been various attempts to devise a notion of emergence which might be usefully applied to purely physical situations (as well as
addition to physical properties, there are physically irreducible domains of emergent properties, of which mental properties are among the leading candidates. Other forms of antireductionism include Davidson's "anomalous monism" and the early functionalism of Hilary Putnam and Jerry Fodor. Property dualism based on ontological physicalism is called nonreductive materialism (or physicalism). Emergentism, anomalous monism, and Putnam-Fodor functionalism are the best-known examples of nonreductive materialism.6

What we have just seen is that this intermediate halfway house between the two poles of substance dualism and reductionist physicalism is a promissory note that cannot be redeemed. As we saw, the supervenience/exclusion argument shows that property dualism is not able to explain how mental causation is possible; instead of saving mental causation, it ends up relegating mental phenomena to the status of epiphenomena. Nonreductive materialism has been motivated by a desire to save mentality as something distinctive and special, and something that we value. Instead of saving it, it loses it by depriving it of causal powers. The important lesson we have learned from three decades of debate—Davidson's "Mental Events" was published in 1970—is this: the demands of causality do not tolerate duality of properties any more than duality of substances, and both Cartesian substance dualism and contemporary property dualism run aground on the rocks of mental causation.

6. The qualifier "Putnam-Fodor" is needed because functionalism of the kind advocated by David Armstrong and David Lewis cannot be viewed as a form of property dualism, at least not in a straightforward way.

So property dualism is out—at least, for now. Where do we go from here? If causality excludes dual realms of mental and physical properties, that means that there is only one secure causal domain, the domain of physical properties. What then happens to mental properties? One possibility is that mental properties are reducible to physical properties; if mental properties are reduced to physical properties, this would conserve and legitimize them as members of the physical domain, thereby safeguarding their causal status. But suppose that the mental fails to reduce: we would then be faced with the specter of epiphenomenalism, and we must find a way to live with causally impotent mental properties. This may very well push us over the edge into mental irrealism; for one might argue that epiphenomenalism is a fate no better than irrealism and in fact indistinguishable from it—Samuel Alexander urged that to deprive something of causal powers is to deprive it of existence.7 This leads to eliminativism: mental properties are banished from our ontology as causally idle "danglers" with no purpose to serve. This is not an outcome that anyone can welcome; most philosophers—for example, Jerry Fodor as we will see below—take mental causation as wholly nonnegotiable; it must be protected at all costs. The best, or the most satisfying, outcome would have been a vindication of mental causation along the lines of nonreductive physicalism; that would have allowed us to retain mentality as something that is causally efficacious and yet autonomous vis-à-vis the physical domain.

But the best outcome, as we saw, is not to be had. The next best outcome, in fact our only hope at this point if mental causation is to be saved, is physical reductionism. Physical reduction would save causal efficacy for mentality, at the cost of its autonomy. Reductionism allows only one domain, the physical domain, but the mental may find a home in that domain. Some

will say that the reductionist option is hardly distinguishable from eliminativism, that to reduce minds and consciousness to patterns of electrical activity in a network of soulless neurons is in effect to renounce them as a distinctive and valued aspect of our being. This reaction is understandable but inappropriate. There is an honest difference between elimination and conservative reduction. Phlogiston was eliminated, not reduced; temperature and heat were reduced, not eliminated. Witches were eliminated, not reduced; the gene has been reduced, not eliminated. We have a tendency to read “nonphysical” when we see the word “mental,” and think “nonmental” when we see the word “physical.” This has the effect of making the idea of physical reduction of the mental a simple verbal contradiction, abetted the idea that physical reduction of something we cherish as a mental item, like thought or feeling, would turn it into something other than what it is. But this would be the case only if by “physical” we meant “nonmental.” We should not prejudge the issue of mind-body reduction by building irreducibility into the meanings of our words. When we consider the question whether the mental can be physically reduced, it is not necessary—even if this could be done—to begin with general definitions of mental and physical; rather, the substantive question that we are asking, or should be asking, is whether or not things like belief, desire, emotion, and consciousness are reducible to neural, biological, and physico-chemical properties and processes. We can understand this question and intelligently debate it, without subsuming these items under some general conception of what it is for something to be mental. If “mental” is understood to imply “nonphysical,” it will then be an open question whether or not belief, desire, sensation, perception, and the rest are mental in that sense. And this question will replace the original question of their physical reducibility. There is here a substantive question which no verbal conventions or decisions should be allowed to trivialize.

In any case, our best remaining option is reductionism. Does this mean that we are committed willy-nilly to reductionism? The answer is no: what we have established, if our considerations have been generally correct, is a conditional thesis, “If mentality is to have a causal influence in the physical domain—in fact, if it is to have any causal efficacy at all—it must be physically reducible.” I have not argued for reductionism simpliciter; rather, I have argued that mental causation requires reduction, and that anyone who believes in mental causation must be prepared to endorse mind-body reduction. We may call this “conditional reductionism.” It is important to keep in mind that this is not reductionism tout court. Moreover, none of this says anything about the truth or plausibility of reductionism. Whether or not the mental can be reduced to a physical base is an independent question that must be settled on its own merits. Those of us who believe in mental causation should hope for a successful reduction. But again this is only a wish; it doesn’t make reducibility real or reductionism true.

So we have finally come to a crossroads: Can we physically reduce minds? Is mentality reducible in physical terms?

Reducing Minds

In raising this question about the reducibility of the mental, it is important not to think that the mental as a totality must be either all reducible or all irreducible. It may well be that parts of the mental are reducible while the rest is not. It may be that

---

8. I have sometimes been described as a reductionist (for example, Robert Van Gulick refers to me as a “hard core reductionist” in his survey article “Reduction, Emergence and Other Recent Options on the Mind-Body Problem: A Philosophical Overview,” Journal of Consciousness Studies 8 (2001): 1–34, p. 2. I am sure I have often written and spoken in confusing ways in the past, but I hope this sets the record straight.