2

The Supervenience Argument Motivated, Clarified, and Defended

An argument was presented in the preceding chapter to show that, on an influential position on the mind-body problem, mental properties turn out to be without causal efficacy. This is what I have called the supervenience argument, also called the exclusion argument in the literature. The argument has drawn comments, criticisms, and objections from a wide range of philosophers, but mostly from those who want to defend orthodox nonreductive physicalism and other forms of mind-body property dualism. Critics of the argument have raised some significant issues, both about the specifics of the argument and, more interestingly, about the broader philosophical issues involved. In this chapter, I would like to address two of the more pressing problems. One is that of "overdetermination," brought up by a number of philosophers; the second is the problem of "causal drainage," forcefully developed by Ned Block in his "Do Causal Powers Drain Away?"[1] Before we get to these and other issues, I want to set out the leading idea that motivates the supervenience argument and then offer what

1. Ned Block, "Do Causal Powers Drain Away?" *Philosophy and Phenomenological Research* 67 (2003): 133–150.

I hope will be a clearer statement of the argument, along with explanatory comments that some may find useful. But first we need a brief description of the philosophical position that is the target of the supervenience argument.

Nonreductive Physicalism

There is no consensus on exactly how nonreductive physicalism is to be formulated, for the simple reason that there is no consensus about either how physicalism is to be formulated or how we should understand reduction. For present purposes, however, no precise formulation is needed; a broad-brush characterization will be sufficient. Moreover, there need not be a single "correct" or "right" formulation of physicalism; there probably are a number of claims, not strictly equivalent, about the fundamentally physical character of the world, each of which can reasonably be considered a statement of physicalism. The strengths and weaknesses, merits and demerits, of these different physicalisms could be examined and debated, and reasonable people could come to different conclusions about them. In any case, most will agree that the following three doctrines are central to nonreductive physicalism: mind-body supervenience, the physical irreducibility of the mental, and the causal efficaciousness of the mental. Mind-body supervenience, the claim that makes the position a form of physicalism, can be stated as follows:

> *Supervenience.* Mental properties strongly supervene on physical/biological properties. That is, if any system *s* instantiates a mental property M at *t*, there necessarily exists a physical property P such that *s* instantiates P at *t*, and necessarily anything instantiating P at any time instantiates M at that time.[2]

2. There are alternative, not quite equivalent, ways of stating mind-body supervenience; one could get a good idea of what these might be from Brian McLaughlin, "Varieties of Supervenience," in *Supervenience: New Essays*, ed. Elias

I take supervenience as an ontological thesis involving the idea of dependence—a sense of dependence that justifies saying that a mental property is instantiated in a given organism at a time *because*, or *in virtue of* the fact that, one of its physical "base" properties is instantiated by the organism at that time. *Supervenience*, therefore, is not a mere claim of covariation between mental and physical properties; it includes a claim of existential dependence of the mental on the physical. I am assuming that a serious physicalist will accept this interpretation of supervenience. Mind-body supervenience as a bare claim about how mental and physical properties covary will be accepted by the double-aspect theorist, the neutral monist, the emergentist, and the epiphenomenalist; it can be accepted even by the substance dualist.

The second component of nonreductive physicalism reflects the "nonreductive" character of this form of physicalism:

> *Irreducibility.* Mental properties are not reducible to, and are not identical with, physical properties.

There is no single well-defined sense, or model, of reduction shared by all disputants in this debate, but this will not matter for us in the context of the supervenience argument; all we need to assume here is that physically irreducible properties remain outside the physical domain—that is, if anything is physically reduced, it must be identical with some physical item. The root meaning of reduction was given, I believe, by J.J.C. Smart when he said that sensations are nothing "over and above" brain processes.[3] If Xs are reduced to Ys, then Xs are nothing over and above the Ys.

Savellos and Ümit Yalçin (Cambridge: Cambridge University Press, 1995). In some contexts the interpretation of "necessarily" as it occurs in the last clause can be crucial; for our purposes, there is no need to opt for any special specification.

3. J.J.C. Smart, "Sensations and Brain Processes," in *The Nature of Mind*, ed. David M. Rosenthal (New York and Oxford: Oxford University Press, 1991), p. 170. Originally published in *Philosophical Review* 68 (1959): 141–56.

We now come to the third doctrine, concerning the causal status of these irreducible mental properties.

> *Causal efficacy.* Mental properties have causal efficacy—that is, their instantiations can, and do, cause other properties, both mental and physical, to be instantiated.

This last thesis is important to the many friends of the position I am describing. The irreducibility claim is often motivated by a desire to save mental properties as something special and distinctive, but if these properties turn out to be causally impotent and explanatorily useless, that would rob them of any real interest or significance, rendering the issue of their reducibility largely moot. Or one could argue that since physical properties are assumed to be causally efficacious, causally inert mental properties obviously cannot be physically reduced. This means that the rejection of mental causal efficacy would make the irreducibility claim true but trivial. In these ways, therefore, the doctrines of irreducibility and causal efficacy go hand in hand.

It can be debated whether these three doctrines constitute a robust enough physicalism. The issue obviously turns on the question whether mind-body supervenience as stated is sufficient for physicalism, since the irreducibility and mental causal efficacy have nothing specifically to do with physicalism; Descartes endorsed both. Moreover, classic emergentism, not usually considered a form of physicalism, endorsed all three, making it a target of the supervenience argument.[4] However, this issue will not affect the discussions to follow. My claims and arguments are intended to apply to any position that accepts the three propositions; what else it accepts makes no difference.

4. See my "Being Realistic about Emergence" in *The Emergence of Emergence*, ed. Paul Davies and Philip Clayton (Oxford: Oxford University Press, forthcoming). The three doctrines, however, can be thought of as capturing the physicalist core of emergentism. On supervenience and physicalism, see Jessica Wilson, "Supervenience-Based Formulations of Physicalism," forthcoming in *Noûs*.

The Fundamental Idea

The idea that drives the supervenience argument can be expressed in the following proposition, which I name after the great eighteenth-century American theologian-philosopher Jonathan Edwards:

> *Edwards's dictum*. There is a tension between "vertical" determination and "horizontal" causation. In fact, vertical determination excludes horizontal causation.

What do I mean by "vertical" determination? Consider an object, say this lump of bronze. At any given time it has a variety of intrinsic properties, like color, shape, texture, density, hardness, electrical conductivity, and so on. Most of us would accept the proposition that the bronze has these properties at this time in virtue of the fact that it has, at this time, a certain microstructure—that is, it is composed of molecules of certain kinds (copper and tin) in a certain specific structural configuration. I describe this situation by saying that the macroproperties of the bronze are vertically determined by its synchronous microstructure. The term "vertical" is meant to reflect the usual practice of picturing micro-macro levels in a vertical array, with the micro underpinning the macro. In contrast, we usually represent diachronic causal relations on a horizontal line, from past (left) to future (right)—"time's arrow" seems always to fly from left to right. From the causal point of view, the piece of bronze has the properties it has at t because it had the properties it had at $t - \Delta t$ (and certain boundary conditions obtained during this period). The past determines the future and the future depends on the past. That is what I mean by "horizontal" causation. So we have here two purported determinative relationships orthogonal to each other: vertical micro-macro mereological determination and horizontal past-to-future causal determination.

The lump of bronze has the color yellow at time t. Why is it yellow at t? There are two presumptive answers: (1) because its

surface has microstructural property M at t; (2) because it was yellow at $t - \Delta t$. To appreciate the force of the supervenience argument it is essential to see a prima facie tension between these two explanations. As long as the lump has microproperty M at t, it's going to be yellow at t, *no matter what happened before t*. Moreover, unless the lump has M, or another appropriate microproperty (with the right reflectance characteristic), at t, it cannot be yellow at t. Anything that happened before t seems irrelevant to the lump's being yellow at t; its having M at t is fully sufficient in itself to make it yellow at t.

So far as I know, Jonathan Edwards was the first philosopher who saw a tension of precisely this kind. Edwards' surprising doctrine that there are no temporally persisting objects was based on his belief that the existence of such objects is excluded by the fact that God is the sustaining cause of the created world at every instant of time. There are no persisting things because at every moment God creates, or recreates, the entire world *ex nihilo*—that is what it means to say that God is the sustaining cause of the world. Consider two successive "time slices" of the bronze: each slice is created by God, and there is no causal or other direct existential relationship between them. To illustrate his argument, Edwards offers a marvelously apt analogy:

> The *images* of things in a glass, as we keep our eye upon them, seem to remain precisely the same, with a continuing, perfect identity. But it is known to be otherwise. Philosophers well know that these images are constantly renewed, by the impression and reflection of *new* rays of light; so that the image impressed by the former rays is constantly vanishing, and a *new* image is impressed by *new* rays every moment, both on the glass and on the eye. . . . And the new images being put on *immediately* or *instantly* do not make them the same, any more than if it were done with the intermission of an *hour* or a *day*. The image that exists at this moment is not at all *derived* from the image which existed at the last preceding moment. As may

> be seen, because if the succession of new *rays* be intercepted, by something interposed between the object and the glass, the image immediately ceases; the *past existence* of the image has no influence to uphold it, so much as for a moment.[5]

Successive images are not causally related to each other; they are each caused by something else. If we suppose that the persistence of an object requires causal relations between its earlier and later stages, Edwards is arguing that "horizontal" causation involving created substances is excluded by their "vertical" dependence on God as a sustaining cause of the world at every instant. Remove God as the sustaining cause; the whole world will vanish at that very instant.[6]

It is simple to see how Edwards's dictum applies to the mind-body case, causing trouble for mental causation. Mind-body supervenience, or the idea that the mental is physically "realized"—in fact, any serious doctrine of mind-body dependence will do—plays the role of vertical determination or dependence, and mental causation, or any "higher-level" causation, is the horizontal causation at issue. The tension between vertical determination and horizontal causation, or the former's threat to preempt and void the latter, has been, at least for me, at the heart of the worries about mental causation.

5. Jonathan Edwards, *Doctrines of Original Sin Defended* (1758), Part IV, Chapter II. The quotation is from *Jonathan Edwards*, ed. C. H. Faust and T. H. Johnson (New York: American Book Co., 1935), p. 335. (Italics in the original.) It seems, however, that Edwards's argument may well have been foreshadowed by the occasionalists of the 17th century.

6. Some will argue that these considerations—and some of the crucial steps in the supervenience argument—depend on the use of a robust, "thick" concept of productive or generative causation rather than a "thin" concept based on the idea of counterfactual dependence or simple Humean "constant conjunctions," and that thin causation is all the causation that there is. See Barry Loewer's "Comments on Jaegwon Kim's *Mind in a Physical World*," *Philosophy and Phenomenological Research* 65 (2002): 655–62, and my reply to Loewer, ibid., 674–77.

The Supervenience Argument Refined and Clarified

Let us now turn to a restatement of the supervenience argument in a more explicit and streamlined form. It is useful to divide the argument into two stages; I believe each stage has its own interest, and this will also enable me to present two materially different ways of completing the second stage of the argument.

Stage 1

We begin with the supposition that there are cases of mental-to-mental causation. Let M and M* be mental properties:

(1) M causes M*.

Properties as such don't enter into causal relations; when we say "M causes M*," that is short for "An instance of M causes an instance of M*" or "An instantiation of M causes M* to instantiate on that occasion." Also for brevity we suppress reference to times. From *Supervenience*, we have:

(2) For some physical property P*; M* has P* as its supervenience base.

As earlier noted, (1) and (2) together give rise to a tension when we consider the question "Why is M* instantiated on this occasion? What is responsible for, and explains, the fact that M* occurs on this occasion?" For there are two seemingly exclusionary answers: (a) "Because M caused M* to instantiate on this occasion," and (b) "Because P*, a supervenience base of M*, is instantiated on this occasion." This of course is where Jonathan Edwards's insight, encapsulated in Edwards's dictum, comes into play: Given that P* is present on this occasion, M* would be there no matter what happened before; as M*'s supervenience base, the instantiation of P* at *t* in and of itself

necessitates M*'s occurrence at *t*. This would be true even if M*'s putative cause, M, had not occurred—*unless*, *that is*, *the occurrence of M had something to do with the occurrence of P* on this occasion*. This last observation points to a simple and natural way of dissipating the tension created by (a) and (b):

(3) M caused M* *by* causing its supervenience base P*.

This completes Stage 1. What the argument has shown at this point is that if *Supervenience* is assumed, mental-to-mental causation entails mental-to-physical causation—or, more generally, that "same-level" causation entails "downward" causation. Given *Supervenience*, it is not possible to have causation in the mental realm without causation that crosses into the physical realm. This result is of some significance; if we accept, as most do, some doctrine of macro-micro supervenience, we can no longer isolate causal relations within levels; any causal relation at level L (higher than the bottom level) entails a cross-level, L to L − 1, causal relation. In short, *level-bound causal autonomy is inconsistent with supervenience or dependence between the levels*. Further, an important part of the interest of the supervenience argument is that it shows that, under the physicalist assumptions we are working with, mind-to-mind causation is in trouble just as much as mind-to-body causation. Often the problem of mental causation is presented as that of explaining how the mental can inject causal influences into the causally closed physical domain, that is, the problem of explaining mental-to-physical causation. I wanted to do something more, namely to show that physicalism can put in peril all forms of mental causation, including mental-to-mental causation.[7] This is why the argument begins with line (1). It is at Stage 2 that we take up mental-to-physical causation. It is noteworthy that,

7. As we will see in the next chapter, an interesting parallel holds in the case of substance dualism: under substance dualism, mental-to-mental causation turns out to be as problematic as mental-to-physical causation.

unlike in the second stage below, the argument up to this point makes no explicit appeal to any special metaphysical principles; in particular, no specific assumptions about the physical domain, such as its causal closure or completeness, enter the picture at this stage.[8] Mental-physical supervenience is the only substantive premise that has been in play thus far.

Stage 2

There are two ways of completing the argument, and I believe the second, which is new, is of some interest. I will first present the original version in a somewhat clearer form:

COMPLETION I

We now turn our attention to M, the supposed mental cause of M*. From *Supervenience*, it follows:

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

There are strong reasons for thinking that P is a cause of P*. I will not rehearse the considerations in support of this idea; let us just note that P is (at least) nomologically sufficient for M, and the occurrence of M on this occasion depends on, and is determined by, the presence of P on this occasion. Since ex hypothesi M is a cause of P*, P would appear amply to qualify as a cause of P* as well. So we have:

(5) M causes P*, and P causes P*.

8. On some occasions I have tried to argue for (3) by invoking an exclusion principle—see, for example, the "principle of determinative/generative exclusion" in chapter 1. I think it preferable not to appeal to any general principle here; I now prefer to rely on the reader's seeing the tension I spoke of in connection with the two answers to the question "Why is M* instantiated on this occasion?" Anyone who understands Jonathan Edwards's argument and his mirror analogy will see it; I don't believe invoking any "principle" will help persuade anyone who is not with me here.

Note that P's causation of P* cannot be thought of as a causal chain with M as an intermediate causal link; one reason is that the P-to-M relation is not a causal relation. Note also that since M supervenes on P, M and P occur precisely at the same time. (Moreover, as we will shortly see, the two principles that will be introduced, *Exclusion* and *Closure*, together disqualify M as a cause of P*, making the idea of a causal chain from P to M to P* a nonstarter.)

To continue, from *Irreducibility*, we have:

(6) M ≠ P.[9]

Again, (5) and (6) present to us a situation with metaphysical tension. For P* is represented here as having two distinct causes, each sufficient for its occurrence. The situation is ripe for the application of the causal exclusion principle, which can be stated as follows:

> *Exclusion*. No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination.

Let us assume that this is not a case of causal overdetermination (we will discuss the overdetermination issue below).

(7) P* is not causally overdetermined by M and P.

By *Exclusion*, therefore, we must eliminate either M or P as P*'s cause. Which one?

9. Note: this only means that this instance of M ≠ this instance of P. Does this mean that a Davidsonian "token identity" suffices here? The answer is no: the relevant sense in which an instance of M = an instance of P requires either property identity M = P or some form of reductive relationship between them. (See *Mind in a Physical World*, ch. 4). The fact that properties M and P must be implicated in the identity, or nonidentity, of M and P instances can be seen from the fact that "An M-instance causes a P-instance" must be understood with the proviso "in virtue of the former being an instance of M and the latter an instance of P."

(8) The putative mental cause, M, is excluded by the physical cause, P. That is, P, not M, is a cause of P*.

We can give relatively informal reasons for choosing P over M as the cause of P*, but for a general theoretical justification we may appeal to the causal closure of the physical domain:

Closure. If a physical event has a cause that occurs at *t*, it has a physical cause that occurs at *t*.[10]

If we were to choose M over P as P*'s cause, *Closure* would kick in again, leading us to posit a physical cause of P*, call it P_1 (what could P_1 be if not P?), and this would again call for the application of *Exclusion*, forcing us to choose between M and P_1 (that is, P). Unless P is chosen and M excluded, we would be off to an unending repetition of the same choice situation; M must be excluded and P retained.

It is worthwhile to reflect on how *Exclusion* and *Closure* work together to yield the epiphenomenalist conclusion (8). *Exclusion* itself is neutral with respect to the mental-physical competition; it says either the mental cause or the physical cause must go, but doesn't favor either over the other. What makes the difference—what introduces an asymmetry into the situation—is *Closure*. It is the causal closure of the physical world that excludes the mental cause, enabling the physical cause to prevail. If the situation with causal closure were the reverse, so that it was the mental domain, not the physical domain, that was causally closed, the mental

10. For discussion of physical causal closure, or "completeness," see, e.g., David Papineau, *Thinking about Consciousness* (Oxford: Clarendon Press, 2002), ch. 1; E. J. Lowe, "Physical Causal Closure and the Invisibility of Mental Causation," in *Physicalism and Mental Causation*, ed. Sven Walter and Heinz-Dieter Heckmann (Exeter, UK: Imprint Academic, 2003). A simpler statement of causal closure in the form "If a physical event has a cause, it has a physical cause" will not do; given the transitivity of causation, the requirement would be met by a causal chain consisting of a physical effect caused by a mental cause which in turn is caused by a physical cause.

cause would have prevailed over its physical competitor. I suppose this could happen under some forms of Idealism; one would then worry about the "problem" of physical causation.

COMPLETION 2

Let us begin with the last line of Stage 1:

(3) M causes M* by causing its physical supervenience base P*.

From which it follows:

(4) M is a cause of P*.

By *Closure* it follows:

(5) P* has a physical cause—call it P—occurring at the time M occurs.

(6) M ≠ P (by *Irreducibility*).

(7) Hence, P* has two distinct causes, M and P, and this is not a case of causal overdetermination.

(8) Hence, by *Exclusion*, either M or P must go.

(9) By *Closure* and *Exclusion*, M must go; P stays.

This is simpler than Completion 1. *Supervenience* is not needed as a premise, and the claim that M's supervenience base P has a valid claim to be a cause of P* has been bypassed, making it unnecessary to devise an argument for it. However, Completion 1, in some ways, is more intuitive; it better captures Jonathan Edwards's fundamental insight and makes it particularly salient how putative higher-level causal relations give way to causal processes at a lower level. Either way, the main significance of Stage 2 lies in what it shows about the possible hazards involved in the idea of "downward" causation, namely that *the assumptions of causal exclusion and lower-level causal closure disallow downward causation*.

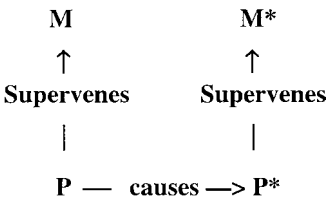

Figure 1.

Figure 1 pictures the outcome of the argument under Completion 1. In this picture, there is but one causal relation, from physical property P to another physical property P*, and the initially posited causal relation from M to M* has been eliminated. An apparent causal relation between the two mental properties is explained away by their respective supervenience on two physical properties that are connected by a genuine causal process. In this picture neither M nor M* is implicated in any causal relations; they play no role in shaping the causal structure—they only supervene on properties that constitute that structure. The supervenience relations together with the causal relation involved can generate counterfactual dependencies between the two mental properties, and between them and the physical properties; but these are no more causal than counterfactual dependencies involving any other supervenient property and its subvenient base (compare the aesthetic properties of a work of art and their base physical properties). Completion 2 presents a picture that is a bit less full: we no longer have the vertical "supervenience" arrow from P to M. M of course must have a physical supervenience base, but the argument, unlike in Completion 1, does not require it to be a cause of P*, although, as Completion 1 suggests, it may well be. The moral, however, is the same: the $M \rightarrow M^*$ and $M \rightarrow P^*$ causal relations have given way to an underlying physical causal process, $P \rightarrow P^*$.

Is Overdetermination an Option?

Several critics have taken issue with line (7), in both completions of the argument, where the claim is made that we should not think of M and P as two distinct overdetermining causes of P*. One thing I said to defend this claim in *Mind in a Physical World* was this: taking the overdetermination option would be in violation of *Closure*, for in a world in which P does not occur but which is as close to the actual world as possible, M would be a cause of P*, leaving P* without a physical cause. My critics have convinced me that what I said there is not quite right and at best incomplete.

Ned Block asks whether in the supposed possible world, one in which the supervenience base P of M does not occur, M could be thought of as occurring at all. If we take away the supervenience base of M, shouldn't that also take away M? This is something to think about. If what Block has in mind is that the following counterfactual may well be true, I agree:

(C) If P had not occurred, M would not have occurred.

For we are apt to reason like this: M was there because P was there, so take away P and M goes as well. "If the patient's nociceptive neurons had not been stimulated at *t*, he would not have experienced pain at *t*," uttered, say, when we deliberately activated these neurons in an experimental situation, would evidently be true. In considering the claim that M and P are each a sufficient cause of P*, however, we need to be able to consider a possible situation in which M occurs without P and evaluate the claim that in this possible situation P* nonetheless follows. If such is not a possible situation—that is, if of necessity any nonP-world is ipso facto a nonM-world—what significance can we attach to the claim that P and M are each an overdetermining sufficient cause of P*, that in addition to P, M also is a sufficient cause of P*? *Supervenience* does not render a nonP-world in

which M occurs impossible; all that *Supervenience* requires is that such a world must include an alternative physical base of M.

So suppose W is a world in which M occurs but P does not. In an instructive and helpful discussion, Thomas Crisp and Ted Warfield have the following to say about such worlds:

> Consider though: either [*Supervenience*] holds in W or it does not. Suppose it does. It follows that M has a physical supervenience base P′ in W. What is the causal status of P′ vis-à-vis P* in W? We won't repeat ourselves, but we saw above an argument of Kim's to the effect that if P′ is a supervenience base for M and M causes P*, then P′ is also causally sufficient for P*. If [*Supervenience*] holds in W, therefore, P* does have a physical cause in W, and [*Closure*] therefore does not fail in W.[11]

Crisp and Warfield are right. Notice, though, that in W, we have a replay of exactly the same situation with which we began—M has a physical base, P′, threatening to preempt it as a cause of P*. In any world in which *Supervenience* holds and M causes P*, some physical property, instantiated at the same time, can claim to be a sufficient cause of P*. As long as *Supervenience* is held constant, there is no world in which M by itself, independently of a physical base, brings about P*; whenever M claims to be a cause of P*, there is some physical property waiting to claim at least an equal causal status. In the actual world, we may suppose that a continuous causal chain connects P with P* (in some cases we may already have detailed neurophysiological knowledge of the physical causal process leading from P to P*).[12] And it would be incoherent to suppose there is another

11. Thomas M. Crisp and Ted A. Warfield, "Kim's Master Argument," *Noûs* 35 (2001): 304–16 (the quoted passage appears on p. 314).

12. In introducing consideration of causal chains, I am implicitly asking the reader to think of causation in terms of actual productive/generative mechanisms involving energy flow, momentum transfer, and the like, and not merely in terms of counterfactual dependencies. Needless to say, the overdetermination idea makes little sense when causation is understood this way.

causal chain from M to P* that is independent of the causal process connecting P with P*; the only plausible supposition is that if there is a causal path from M to P*, that must coincide with the causal path from P to P*. In W, another causal chain connects P′ with P*, and the M-P* chain must coincide with that, and similarly in other such worlds. To be a cause of P*, M must somehow ride piggyback on physical causal chains—distinct ones depending on which physical property subserves M on a given occasion, in the same world or in other possible worlds. And we may ask: In virtue of what relation it bears to physical property P does M earn its entitlement to a free ride on the causal chain from P to P* and to claim this causal chain to be its own? Obviously, the only significant relation M bears to P is supervenience. But why should supervenience confer this right on M? The fact of the matter is that there is only one causal process here, from P to P*,[13] and M's supposed causal contribution to the production of P* is totally mysterious. In standard cases of overdetermination, like two bullets hitting the victim's heart at the same time, the short circuit and the overturned lantern causing a house fire, and so on, each overdetermining cause plays a distinct and distinctive causal role. The usual notion of overdetermination involves two or more separate and independent causal chains intersecting at a common effect. Because of *Supervenience*, however, that is not the kind of situation we have here. In this sense, this is not a case of genuine causal overdetermination, and *Exclusion* applies in a straightforward way. Moreover, anyone tempted by the idea that mental events make their causal contributions by being

13. Some have suggested that the M-to-P* causation is a higher-level "redescription" of the causal process from P to P*. E.g., John R. Searle, "Consciousness, the Brain and the Connection Principle: A Reply," *Philosophy and Phenomenological Research* 55 (1995): 217–32, especially 218–19. Obviously, the redescription strategy is available only to those who accept "M = P," namely reductionist physicalists (Searle of course does not count himself among them).

overdetermining causes should reflect on whether this option could sufficiently vindicate the causal efficacy of the mental.

Now for the second leg of Crisp and Warfield's dilemma:

> Now suppose that [*Supervenience*] does not hold in W. And suppose further that, just as Kim suggests, M causes P* in W without there being any physical cause of P*. Given these assumptions, [*Closure*] does indeed fail in W. But recall that we have supposed along with Kim that the actual world is a Supervenience-world. It follows from this supposition that W is either nomologically or metaphysically impossible, depending on how we read the relevant modal operator in the formulation of [*Supervenience*]. So if W is a world in which [*Closure*] is violated in the way Kim suggests, W is at least nomologically impossible.
>
> What should nonreductivist fans of overdetermination think about this? Should they give up their view because it implies that [*Closure*] fails in worlds that are nomologically (and maybe even metaphysically) impossible? We can't see why they should.[14]

I think we can set aside the possibility that mind-body supervenience is logically or metaphysically necessary, since such a view is essentially a reductionist view,[15] and we are here considering *Supervenience* as a part of nonreductive physicalism. Let us assume then that *Supervenience* is nomologically necessary, and that it fails in W. So in virtue of violating *Supervenience*, W is nomologically impossible. However, W is nomologically impossible not because some physical law is violated in W but because some mental properties fail to supervene on physical properties—that is, because some psychophysical laws of our world fail in W. So W may well be a physically possible world; in fact, we may stipulate W to be a perfect duplicate of our

14. Crisp and Warfield, "Kim's Master Argument," p. 314.

15. This is not an uncontroversial issue, but we cannot go into it here. And there are independent reasons for thinking that mind-brain supervenience, if it holds, must be construed as nomological, not logical or metaphysical, supervenience.

world in all physical respects, including spacetime structure, basic physical laws, and fundamental particles. Should the physicalist not care whether physical causal closure holds in a world like W? Contrary to what Crisp and Warfield suggest, it seems obvious to me that anyone who cares about physicalism should care very much about *Closure* in W.

A more direct way of ruling out overdetermination as an option is to adopt a stronger form of physical causal closure:

> *Strong closure*. Any cause of a physical event is itself a physical event—that is, no nonphysical event can be a cause of a physical event.[16]

Using this principle as a premise has two significant effects. First, it stops the overdetermination option in its tracks; *Strong closure* by itself disallows mental-to-physical causation. Second, *Strong closure* allows us to dispense with *Exclusion*. We no longer need this principle to exclude M in favor of P as P*'s cause, for the simple reason that *Strong closure*, in conjunction with *Irreducibility*, makes M ineligible as a cause of P*.

How might the supervenience argument go under *Strong closure*? Stage 1 is unaffected. Let's briefly look at how Completion 1 might go with *Strong closure*:

(3) M causes M* by causing P*.

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

(5) M causes P*, and P causes P*.

Up to here, the argument is the same as before; from here the argument can continue as follows:

(6*) For every physical property P, M ≠ P *Irreducibility*.

(7*) M does not cause P* (from (6*) and *Strong closure*).

16. An even stronger form of closure can be obtained by also prohibiting physical events from having mental effects—that is, by disallowing all "mixed" causal chains, chains with both physical and mental events.

(8*) M does not cause M* (from (3)[17] and (7*)).

(9*) P causes P* (from (5)).

The outcome is the same as in the original Completion 1, namely Figure 1. But the argument has been simplified in that *Exclusion* has been dispensed with as a premise.

Is this a reason to prefer *Strong closure* to *Closure*? The answer, I believe, is yes and no. Although the causal exclusion principle has been widely accepted and I believe it is virtually an analytic truth with not much content, some find it problematic, and the fact that *Strong closure* makes *Exclusion* dispensable is a point in its favor. (This need not be taken to mean that the argument is no longer properly called an "exclusion" argument; even though no exclusion principle is used as a premise, the *outcome* of the argument is that mental causal relations are "excluded" by physical causality.) Further, there seems no reason for the physicalist to object to *Strong closure*; so why not trade the two premises, *Closure* and *Exclusion*, for a single premise, *Strong closure*, and in the process defuse the overdetermination issue? I believe, though, that there is a philosophical gain in staying with the weaker closure premise. Adopting *Strong closure* as a premise is like starting your argument with mind-body causation already ruled out, at least for nonreductivists; with *Strong closure* as your starting point, there isn't very much more distance you can go or need to go. Perhaps philosophical arguments never make converts out of those who are already committed to the opposite side; but I believe that it can serve philosophical interest to begin with a set of premises that are individually as weak as possible but which somehow conspire together to yield the desired conclusion. It is better, that is to say, to distribute the burden of defending a conclusion among a set of relatively weak premises than to place it on fewer but individually stronger premises.

17. It is implicit in (3) that this is the *only* way M can cause M*.

The latter strategy is apt to provoke the complaint that the argument begs the question and that it serves no useful purpose. I think we learn something about the issues and desiderata involved and their interplay when we run the supervenience argument with *Closure* rather than *Strong closure*.

The Generalization Argument

My main aim in this chapter is to respond to the argument Block has put forward in the following passage:

> The Exclusion Principle [the thesis that "sufficient causation at one level excludes sufficient causation at another level"] leads to problems about causal powers draining away. Kim discusses a number of such problems, including the following two. First, it is hard to believe that there is no mental causation, no physiological causation, no molecular causation, no atomic causation but only bottom level physical causation. Second, it is hard to believe that there is no causation at all if there is no bottom level of physics.[18]

Why does Block think that if the supervenience argument holds, there will be no physiological causation, no molecular causation, etc. any more than mental causation? Because he subscribes to what is called the "generalization argument"—the idea that the supervenience argument generalizes beyond mind-body causation, with the result that causation at *any* level gives way to causation at the next lower level (if there is one), just as the supposed causation at the mental level gets eliminated in favor of causation at the physical/biological level. Block is not alone here. A number of writers have expressed the view that if the supposed problem of mental causation is a real problem, a parallel problem should arise for all other special

18. Block, "Do Causal Powers Drain Away?" p. 138.

sciences, except causation at the most fundamental physical level.[19] Such a view is often stated against the backdrop of a "layered" model of the domains of science, according to which objects and properties of the world are arrayed in a hierarchy of "levels," with the basic physical particles and their properties at the bottom level and, above it, the levels of atoms, molecules, cells, organisms, and so on, all ordered in an ascending ladder-like structure. It is this hierarchical view of the domains of science that gives meaning to the talk of "higher" and "lower" levels—in regard to sciences, laws, explanations, and the rest.[20]

On a hierarchical picture of levels like this, it is natural to think of mental causation only as a special case of higher-level causation. If the supervenience argument shows causation at the psychological level to be preempted by causation at the biological level, why couldn't the argument be iterated to show biological causation to be preempted by physicochemical causation, and so on down to the fundamental microphysical level? The idea that the argument is generalizable this way gains force from the widely accepted assumption that properties at upper levels are supervenient on lower-level properties, the eponymous premise that plays a crucial role in the argument.

Let me begin my response by pointing out that if indeed the supervenience argument is generalizable, that only shows that

19. This includes Tyler Burge, Robert Van Gulick, and many others. See my *Mind in a Physical World*, ch. 3 for references and discussion. Among other discussions of the generalization argument are Paul Noordhof, "Micro-Based Properties and the Supervenience Argument," *Proceedings of the Aristotelian Society* 99 (1999): 109–114; Carl Gillett, "Does the Argument from Realization Generalize? Responses to Kim," *Southern Journal of Philosophy* 39 (2001): 79–98; Thomas D. Bontly, "The Supervenience Argument Generalizes," *Philosophical Studies* 109 (2002): 75–96.

20. Whether a layered model of this kind can be developed as a comprehensive ontology of the world is a debatable issue. I discuss some of the difficulties with such an approach in "The Layered Model: Metaphysical Considerations," *Philosophical Explorations* 5 (2002): 2–20. See also John Heil, *From an Ontological Point of View* (Oxford: Oxford University Press, 2003), ch. 4.

we have a general philosophical problem on hand, and that it is not necessarily a refutation of the argument. If the argument goes wrong, one would like to know just where and how it goes wrong. Moreover, just saying that there "obviously" are biological causation, physiological causation, and so on isn't very helpful; what has to be shown is that these kinds of "higher-level" causation are irreducible to basic physical causation—namely, that there are these causal relations *in addition to* the underlying physical causal processes. It is important to keep in mind that the supervenience argument assumes among its premises the doctrine of the irreducibility of the mental; this premise is invoked at line (6) in both completions of Stage 2. As may be recalled, the argument begins with the supposition that an instance of a mental property M causes another mental property M* to instantiate (line (1)). Block says that this M-to-M* causal relation is "putative—it is a premise in a *reductio* that Kim will reject."[21] But this is not the full story: there is another premise, the premise of irreducibility (line (6): $M \neq P$), against which a reductio can also be performed. This premise, not the supposed M-to-M* causal relation, has always been my primary target. The real aim of the argument, as far as my own philosophical interests are concerned, is not to show that mentality is epiphenomenal, or that mental causal relations are eliminated by physical causal relations; it is rather to show "either reduction or causal impotence." To put it another way, my aim is to force a choice between the situation depicted in figure 1 and what is pictured in figure 2. In this picture, the $M \rightarrow M^*$ causation remains genuine and real; it is the very same causal relation as $P \rightarrow P^*$; the reduction collapses the two levels into one, and there is here one causal relation, not two. The aim of the supervenience argument is to clarify the options available to the physicalist: If you deem yourself a

21. Block, "Do Causal Powers Drain Away?" p. 134.

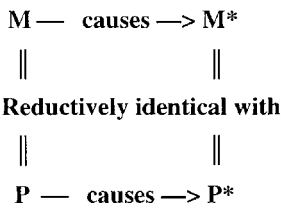

Figure 2.

physicalist, you must choose between figure 1 and figure 2. There are no other options.[22]

Indeed, the supervenience argument may be generalizable, but all that would show is that if there is biological causation, biological properties are, or are reducible to, physical or physicochemical properties; it does not show that biological causation does not exist. The epiphenomenalist brunt of the argument is avoided if one is prepared, and is able, to choose the reductionist branch of the dilemma. It should be kept in mind that merely "choosing" reductionism doesn't make reductionism true; whether or not reductionism is sustainable as an option is an independent question that ought to be decided on its merits.

Many philosophers will reply that biological properties are no more physically reducible than psychological properties, citing their "multiple realizability" in relation to physicochemical properties. For most antireductionist philosophers, multiple realizability has long been a mantra, an all-purpose antireductionist argument applied across the board to all special science properties. They see multiple realization everywhere, and this

22. The underlying metaphysical moral of the two options is the same, however: there is only one causal relation here, namely a physical one, and, more generally, causality is fundamentally a physical phenomenon. An interestingly similar picture results from Donald Davidson's thesis that causation requires "strict laws," and that strict laws are found only in physics. See his "Mental Events," in *Essays on Actions and Events* (Oxford and New York: Oxford University Press, 1980).

leads them to see irreducibility everywhere. I believe, however, that the notion of "realization" as it is often invoked in this context is too loose and ill-formed, and that when realization is properly understood, multiple realization only leads to reducibility to multiple reduction bases, not to irreducibility.[23]

Considerations like those motivating the supervenience argument do not have eliminative implications for macrocausation in general; the supervenience argument does not eliminate all macrocausation, leaving only causal relations between microentities and their properties. This baseball has causal powers that none of its proper parts, in particular none of its constituent microparticles, have, and in virtue of its mass and hardness, the baseball can break a window when it strikes it with a certain velocity. The shattering of the glass was caused by the baseball and certainly not by the individual particles composing it. True, the baseball is a composite object made up of its constituent molecules, atoms, particles, or what have you, and this complex structure consisting of microparticles broke the window. But there is no mystery here: the baseball = this composite structure of microparticles.[24] Presumably, the causal powers of the baseball are *determined* by its microstructural features and perhaps also *explainable* in terms of them. But determination or explanation need have no eliminative implications. Perhaps, macrocausal relations are constituted by, or composed of, a bunch of microcausal relations. But that does not banish macrocausation out of existence any more than the fact that the baseball is composed of microparticles entails its nonexistence. All this is consistent with the supervenience argument.

23. See my "Multiple Realization and the Metaphysics of Reduction," *Philosophy and Phenomenological Research* 52 (1992): 1–26, and *Mind in a Physical World*, ch. 4. For further discussion of multiple realizability and reduction, see John Bickle, *Psychoneural Reduction: The New Wave* (Cambridge, MA: MIT Press, 1998).

24. For a dissenting view—*plus* the view that macrocausation is in general preempted by microcausation—see Trenton Merricks, *Objects and Persons* (Oxford: Clarendon, 2001).

Block's Causal Drainage Argument

A micro-based property of an object is a property characterizing its microstructure—it tells us what sorts of micro-constituents the object is made up of and the structural relations that configure these constituents into a stable object with substantival unity. Micro-based (or microstructural) properties of an object are its macroproperties—they belong to the whole object, not to its constituents—and, moreover, they do not supervene on the properties of the object's micro-constituents. For that reason, the supervenience argument does not touch micro-based properties,[25] and I have claimed that this prevents causal powers from seeping downward from level to level, from macro to micro. Further, I have argued that many chemical and biological properties seem construable as micro-based properties, properties defined or analyzable in terms of microstructure. Block recognizes this as my strategy. The initial criticism he advances can be called the "multiple composition" argument. He writes:

> But why can't micro-based properties be micro-based in *alternative ways*? Why isn't jade an example of a micro-based property, micro-based in both calcium magnesium silicate (nephrite) and sodium aluminum silicate (jadeite)?. . .
>
> My doubts about [Kim's] picture center on the worry just mentioned concerning multiple decomposition. Micro-based properties are supposed to prevent draining away for both supervenient and functional properties, but Kim's plugging the draining with micro-based properties depends on assuming identities (such as "water = H_2O") and multiple composition will preclude such identities.[26]

25. This has been disputed by some of the authors cited in footnote 19.
26. Block, "Do Causal Powers Drain Away?" pp. 145–46.

Here Block appears to be thinking of multiple composition in parallel with multiple realization: just as multiple realization has been used as an argument against reducibility, multiple composition could be used against identifying a macroproperty, say being jade, with micro-based properties. This is an interesting possibility; multiple compositionality may work as well as multiple realizability, each against its reductionist target. However, I think that neither works very well.

There are two things to say about Block's argument. First, in spite of jade's multiple composition, each instance of jade—that is, each individual piece of jade—is either jadeite or nephrite, and I don't see anything wrong about identifying *its* being jade with *its* being nephrite (if it is nephrite) or with *its* being jadeite (if it's jadeite). If it is nephrite, the causal powers that it has in virtue of being jade will be exactly identical with the causal powers of nephrite. All we need is identity at the level of instances, not necessarily at the level of kinds and properties; causation after all is a relation between property or kind-instances, not between properties or kinds as such. Second, suppose a macroproperty has two or more distinct micro-compositions. We can use the jade example again: we presumably distinguish between the two compositions, jadeite and nephrite, importantly because they are *causally* distinguishable—that is, jadeite and nephrite have significantly different causal profiles. Given this, there are two options. We can either deny that jade is a genuine kind (at least, jade is not a kind of mineral), on account of its causal heterogeneity, or identify jade with a disjunctive kind, jadeite or nephrite (that is, being jade is identified with having the microstructure of jadeite or the microstructure of nephrite). The second option which allows disjunctive kinds is a more conservative approach and may be more viable as a general solution. On the disjunctive approach, being jade turns out to be a causally heterogeneous property, not a causally inert one, and jade turns out to be a causally heterogeneous kind, not a causally irrelevant one. To disarm

Block's multiple composition argument, adopting either disjunctive property/kind identities or instance (or token) identities seems sufficient.

This, however, does not fully block the drainage argument. There may be no causal seepage from macro to micro, but that is not the only way the seepage can occur. The trouble can be seen when we recognize that a given object can have micro-based properties at various levels (the biological, the physico-chemical, the atomic, etc.), and that higher-level micro-based properties arguably supervene on their lower-level counterparts. Block has this in mind, I think, when he speaks of "endless subvenience."[27] Other commentators, in particular Ausonio Marras,[28] have also made this point. Let us see how the idea might be developed.

Take any macro-object, O, and let a *total* micro-based property *at level L* be the property corresponding to a complete description of O's microstructure at level L. (Roughly, we can think of "levels" in terms of modes of decomposition of material objects into physically significant constituents; examples of levels are the molecular level, the atomic level, and the level of basic particles.) So if L is the level of the Standard Model, a total micro-based property of O at this level would give a complete description of O's microstructure in terms of the particles and forces posited in the Standard Model. The following is a plausible physicalist principle:

> *Macro-micro supervenience*. All intrinsic properties of O, at any level higher than L, supervene on the total micro-based property of O at level L.

The idea is that wholes made up of the same (qualitatively identical) constituents configured in the same structural relationships

27. Block, "Do Causal Powers Drain Away?" p. 140.

28. Ausonio Marras in "Critical Notice of *Mind in a Physical World*," *Canadian Journal of Philosophy* 30 (2000): 137–60; see p. 151.

will exhibit an identical set of intrinsic properties. Since micro-based properties are intrinsic properties, it follows:

> For any object O, O's micro-based properties at level L supervene on O's total micro-based property at level L*, where $L^* < L$.

Consider a series of total micro-based properties of a given object: M_L, M_{L-1}, M_{L-2}, ... Suppose this series has no end; it continues on, without ever reaching a bottom level. That is, let us suppose that the speculation of the physicists cited by Block is correct, and that matter is infinitely divisible (I will go along with Block that all this makes perfectly good sense; but can we really make sense of the idea of an object that is literally made up of infinitely many physically significant parts, here and now?) According to the supervenience argument, M_L apparently cedes its causal powers to M_{L-1}, whose causal powers in turn are taken over by those of M_{L-2}, and so on without end.

Here, Block's worry appears well placed. The supervenience argument implies the following general proposition:

> *Seepage*. If property Q supervenes on a property Q* at a lower level without being reducible to it, Q's causal powers are preempted by those of Q*.

This means that no member of the infinite series of total micro-based properties M_L, M_{L-1}, . . . has causal powers, since every member has a lower member on which it supervenes. If no member of this series has causal powers, there are none to be had anywhere in the series. Moreover, since all intrinsic properties of the object in question are assumed to supervene on its total micro-based properties at lower levels, none of the object's intrinsic properties can have causal powers, and that means that the object itself has no causal powers. All this on the premise that microphysics has no bottom level and matter is infinitely divisible.[29]

29. For an interesting (skeptical) discussion of the existence of a bottom level, see Jonathan Schaffer, "Is There a Fundamental Level?" *Noûs* 37 (2003): 498–517.

This, I believe, is Block's argument, or at least it is a close-enough approximation to it. As Marras has pointed out, it seems possible to develop the generalization argument within a single level in the micro-macro hierarchy. In any case, the argument is worth thinking about. Compare *Seepage* with the following alternative ways of conceiving the interlevel causal relationship:

> *Explanation*. If property Q supervenes on a property Q* at a lower level without being reducible to it, Q's causal powers (and the causal relations into which Q enters) can be *explained* in terms of the causal powers of Q*.
>
> *Constitution*. If Q supervenes on Q*, Q's causal powers are *constituted* by those of Q*.
>
> *Derivation/determination*. If Q supervenes on Q*, Q's causal powers *derive from*, and are *determined by* and *dependent on*, those of Q*.

It is interesting to note that, unlike *Seepage*, none of these alternatives seem to be vulnerable to the drainage argument. The reason is that these alternatives, insofar as we understand them, don't appear to have eliminative implications for causation at the higher, supervenient levels. For example, the fact that Q's causal powers are "explained" by the causal powers of its underlying base Q* does not mean that the former are in any sense preempted or eliminated by the latter, or even that they are somehow reduced to the latter. Exactly what "constitution"[30] might mean, or what "derivation" and "dependence" amount to, requires further thought, but it is clear that these

30. For a defense of nonreductive physicalism based on the idea of constitution, see Derk Pereboom, "Robust Nonreductive Physicalism," *Journal of Philosophy* 99 (2002): 499–531. I believe that the main burden, which is yet to be discharged, of this approach is to produce a serviceably clear concept of constitution. See also Lynne Rudder Baker, *Persons and Bodies: A Constitution View* (Cambridge: Cambridge University Press, 2000).

terms as understood in their rough ordinary philosophical senses have no obviously eliminative intimations.

So why not embrace one or another, or perhaps a combination, of these alternative ways of conceiving the interlevel causal relationships? That would stop the drainage right at the start, and whether there is, or is not, a bottom level makes no difference. So why not say that M, though it doesn't quite have the causal status of P in relation to P*, is a "derivative" cause of P* in virtue of its supervenience on P? M is not in itself an independent cause of P*; its causal status derives from its supervenience on the causally active P. Some years back, I thought that this might be a plausible way of vindicating mental causation.[31] This was the model of so-called supervenient causation. But it soon began to dawn on me that this was an empty verbal ploy; we can "say," if we want, that M is a "supervenient" cause, "dependent" or "derivative" cause, or whatever, and we can embellish *figure 1* by drawing a horizontal arrow connecting M with M*, with the annotation "superveniently causes," as in figure 3. But this is only a gimmick with no meaning; the facts are as represented in the unadorned figure 1, and inserting a dotted arrow and calling it "supervenient" causation, or anything else (how about "pretend" or "faux" causation), does not alter the situation one bit. It neither adds any new facts nor reveals any hitherto unnoticed relationships. Inserting the extra arrow is not only pointless; it could also be philosophically pernicious if it should mislead us into thinking that we have thereby conferred on M, the mental event, some real causal role. Moreover, embracing this approach would lead us back to the overdetermination/exclusion problem—unless we simply stipulate the problem away by declaring that supervenient causal relations do not compete with the causal relation underlying them.

31. In, e.g., "Epiphenomenal and Supervenient Causation," *Midwest Studies in Philosophy* 9 (1984): 257–270. See also Ernest Sosa, "Mind-Body Interaction and Supervenient Causation," *Midwest Studies in Philosophy* 9 (1984): 271–81.

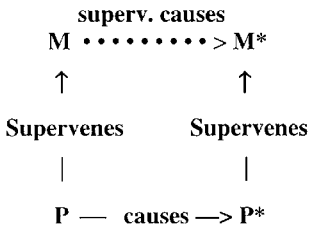

Figure 3.

Jonathan Edwards would have approved my position: in his argument against persisting objects, he did not settle for derivative or dependent causation between created substances; he felt, rightly, that God wholly preempted causation at "higher" levels.

What is Block's own position in regard to these issues? He writes:

> But there is another point of view that recognizes causal efficacy at many levels and does not regard them as competing. And this latter point of view also avoids the problem of causal powers draining away.[32]

And, concerning the "tension" I described at steps (1) and (2) of the supervenience argument, Block writes: "Of course, the non-reductive materialist who accepts causation at many levels should not recognize any tension."[33]

Block's position is the favored approach of most nonreductive physicalists; however, this popular position is precisely what is being challenged. The nonreductive physicalist who accepts supervenience *ought to* recognize the tension; in view of

32. Block, "Do Causal Powers Drain Away?" p. 149. Terence Horgan, among others, holds a similar view; see his "Nonreductive Physicalism and the Explanatory Autonomy of Psychology," in *Naturalism: A Critical Appraisal*, ed. Stephen J. Wagner and Richard Warner (Notre Dame: IN: Notre Dame University Press, 1993).

33. Block, "Do Causal Powers Drain Away?" p. 135.

the considerations advanced in the supervenience argument—basically, Jonathan Edwards's insight—I believe the nonreductive physicalist owes us an explanation of why there is no tension here. It would be nice if we could embrace causation at many levels, including the psychological, the biological, and so on, and also cross-level causation, both downward and upward, all of them coexisting in harmony. And it *is* important to us to be able to have trust in the causal efficacy of our beliefs and desires, emotions and consciousness, and to believe in our powers as agents in the world—all this without reducing mentality to mere patterns of electrical activity in the brain. But these are only a wish list—the starting point of the mental causation debate. The main purpose of the supervenience argument is to bring into focus the disquieting fact that there are strong metaphysical pressures on our pre-philosophical assumptions and desiderata in this area. If the argument is correct, it shows that there are inevitable causal entanglements between different levels, raising all sorts of issues concerning causal closure, competition, and exclusion, and forcing some significant philosophical choices. The nonreductive materialist must sort out and come to terms with these issues; ignoring them is not an option for him. With his drainage argument, Block attempts to defeat the supervenience argument. That is a first step. But this argument has the form of a reductio: if it works, we will know the argument cannot be sound, but that will not tell us just where the argument goes wrong. And this knowledge is required if we are to construct a positive account of multilevel causation of the sort that Block and others have in mind.

In any case, what can be said to counter the drainage argument as lately formulated? As far as the dialectics of the mental causation debate goes, my response here is the same as my reply to Block's statement that the supervenience argument is a reductio against its first premise ("Mental property M causes mental property M*"). As may be recalled, I pointed out that there is another premise against which a reductio can be

performed, namely the premise of psychophysical irreducibility, and that this was my real target. If, as Block's argument suggests, the supervenience argument can be continued to yield as a further conclusion the following proposition:

(H) If there is no bottom level in microphysics, there is no causation anywhere.

and if we find (H) unacceptable, that only means that we need to consider which of the premises of the argument is to be rejected. My suggestion again is that the irreducibility premise should be the prime candidate for rejection; I will elaborate on this below.

Before we go on, there is one point that needs to be clarified. Contrary to what seems sometimes assumed, it is not the case that according to my argument, causation at any level L gives way to causation at level L − 1 (the next lower level), like the rungs of a ladder that keep collapsing each on top of the next lower one. That this is not the case is seen from the fact that the argument requires *Closure* as a premise—the assumption that the lower level in play is causally closed. This means that the mental rung will not collapse onto the biological rung, as far as the supervenience argument is concerned, for the simple reason that the biological level is not causally closed. The same is true of macrolevel physics and chemistry. It is only when we reach the fundamental level of microphysics that we are likely to get a causally closed domain.[34] As I understand it,

34. Actually various complications arise with the talk of levels in this context. In the only levels scheme that has been worked out with some precision, the hierarchical scheme of Paul Oppenheim and Hilary Putnam (in their "Unity of Science as a Working Hypothesis," *Minnesota Studies in the Philosophy of Science*, vol. 2, Minneapolis: University of Minnesota Press, 1958), it is required that each level includes all mereological aggregates of entities at that level (that is, each level is closed under mereological summation). Thus, the bottom level of elementary particles, in this scheme, is in effect the universal domain that includes molecules, organisms, and the rest.

the so-called Standard Model is currently taken to represent the bottom level. Assume that this level is causally closed; the supervenience argument, if it works, shows that mental causal relations give way to causal relations at this microlevel. And similarly for biological causation, chemical causation, geological causation, and the rest. So as far as the supervenience argument goes, the bottom level of fundamental particles (assuming that this is the only level that is causally closed) is always the reference physical domain; there is no step-by-step devolution of causal relations from level to level (I am not suggesting that Block thinks that).[35]

Block's drainage argument evokes some deep metaphysical associations, and this is part of what makes it so interesting. Just think of the whole group of celebrated philosophical arguments with a similar structure, going back to Aristotle and Aquinas. I have in mind Aristotle's argument for the existence of a "prime mover"—the unmoved mover that is the source of all motion. If something moves, it is moved by another thing that moves, which in turn is moved by yet another mover, and so on; but this series cannot go on ad infinitum, for that would make motion impossible. So there must be a mover that is itself not moved by anything else. Aquinas's cosmological argument for the existence of God appears to work in a similar way: there must be a first cause that is itself uncaused because the causal series cannot extend into the past without end. If it did, nothing would exist. The classic foundationalist argument, such as we find in Chisholm,[36] for the existence of "basic" knowledge runs the same way, as do the familiar arguments for

35. A similar problem may well arise for mind-body supervenience; it is likely that mental properties do not supervene on biological properties alone, and that to get full supervenience we have to reach further down and include nonbiological physicochemical properties in the base.

36. Roderick M. Chisholm, *Theory of Knowledge*, 2nd edition (Englewood Cliffs, NJ: Prentice-Hall, 1977).

the existence of semantic primitives, the existence of intrinsic goods, and the like. I think it would be interesting to analyze the metaphysics and logic of arguments that share this general structure. Here, however, I will only make a couple of points specifically in regard to Block's drainage argument.

The first point concerns causal closure. As earlier noted, a causal collapse to the level below would occur only if the lower level is causally closed. Are we assuming that if matter is infinitely divisible, physics will be causally closed at each level of decomposition? I believe that the physicist David Bohm made the observation that each time we descend to a lower microlevel, we do so because the current level is not causally closed ("explanatorily complete" may be a better term in this context); that is, because there are phenomena at this level that can only be explained by descending to a lower level. If something like that is true, no level in Block's infinitely descending series of levels will be causally closed, or explanatorily complete, and the supervenience argument cannot get a toehold. We would not have the required closure premise available—unless we take as our lowest level the *union* of all the microlevels in this infinite chain. Will such a union be causally closed? It has to be, and I believe it may well give us the bottom level which will stop Block's infinite causal drainage.

Second, we must return to reduction again. For Block's drainage argument to work in full force, it must be assumed that the irreducibility premise will hold for purely physical levels—we must assume that molecular facts are not reducible to atomic facts, that atomic facts are not reducible to facts at the level of the Standard Model, and so on down the line. How plausible is this assumption? There are well-known, though by no means undisputed, arguments for regarding the mental to be physically irreducible, and arguments have been advanced to show that the biological level is irreducible to the physicochemical level. But I know of no argument, other than Block's multiple-composition argument discussed above, to

show that the irreducibility assumption will stand as we go down from one microphysical level to the next. The standard view, as I understand it, is that chemistry and macrophysics are reducible, and in fact have already been substantially reduced, to particle physics via quantum mechanics.[37] Unless we have reason to think that irreducibility will hold "all the way down," we have no reason to think that the causal drainage will go on forever. Reduction is the stopper that will plug the cosmic hole through which causal powers might drain away.

In fact, there appear to be presumptive reasons for thinking that reducibility will hold for the kind of infinite series Block has in mind. Let us begin by noting that in various philosophical contexts the identity "the property of being water = the property of being H_2O" is often affirmed and accepted. This identity is accepted presumably on the basis of the fact that water = H_2O. Let us think a bit about what is involved. The property of being H_2O is a total micro-based property of water at the atomic/molecular level; it is the property of being made up of two hydrogen atoms and one oxygen atom in a certain relational structure. Being water is having this kind of microstructure. Having this microstructure is the microstructural essence of water, and being water just is having that structure. We must expect this line of thought to generalize downward, and the following may be one way to flesh it out. Let us say that the property of being H_2O is the total micro-based property of water at the atomic level L (so having M_L = being H_2O). So we have:

(1) Being water = having M_L.

At the next level down, L−1, say the level of the Standard Model, hydrogen atoms have a certain microstructural composition as do oxygen atoms, and water has a certain microstructural

37. See, e.g., Brian P. McLaughlin, "The Rise and Fall of British Emergentism," in *Emergence or Reduction*? ed. Ansgar Beckermann, Hans Flohr, and Jaegwon Kim (Berlin: De Gruyter, 1992).

composition at this level; call it M_{L-1}. Then by the same reasoning that led us to (1), we have:

(2) Being water = having M_{L-1}.

At the level L−2, the one below the Standard Model (if there is such a level), water is again going to have a certain microstructure at that level; this is M_{L-2}. We then have:

(3) Being water = having M_{L-2}.

and so on down the line, to M_{L-3} and the rest. These identities in turn imply the following series of identities:

$M_L = M_{L-1} = M_{L-2} = M_{L-3} \ldots .$

Voilà! These are the identities we need to stop the drainage.

The foregoing is somewhat sketchy and perhaps too quick, and I do not wish to rest my reply to Block's drainage challenge wholly on these rather speculative thoughts. The primary response to the drainage argument is the point that for downward causal drainage to occur, the reduction option must be ruled out for purely physical levels, including microphysical levels, and it is far from obvious that this can be done. In fact, the drainage problem provides us with one more reason to perform a reductio against the irreducibility premise of the supervenience/exclusion argument.

3

The Rejection of Immaterial Minds

A CAUSAL ARGUMENT

The deep difficulties that beset contemporary nonreductive physicalism might prompt some of us to explore nonphysicalist alternatives; in fact, the nonreductivist's predicament seems to have injected new vigor into the dualist projects of philosophers with antecedent antiphysicalist sympathies.[1] For the upshot of our considerations on mental causation was that, for the physicalist, there are only two options left: reductionism and epiphenomenalism. With good reason, most philosophers have found neither choice palatable. On one hand, epiphenomenalism strikes most of us as obviously wrong, if not incoherent; the idea that our thoughts, wants, and intentions might lack causal efficacy of any kind is deeply troubling, going as it does against everything we believe about ourselves as agents and cognizers. It is the kind of doctrine—perhaps radical skepticism is another example—that, even if we had to acknowledge it as true, could not serve as a guide to life; it cannot serve as a premise in our practical reasoning, and it is not possible for us

1. For example, William Hasker, *The Emergent Self* (Ithaca, NY: Cornell University Press, 1999); Timothy O'Connor, *Persons and Causes* (Oxford and New York: Oxford University Press, 2001).

to live as though it is true.[2] Reductionism, on the other hand, has seemed to many people not much better: if minds turn out to be mere configurations of neurons, silicon chips, or whatever and consciousness and thoughts are simply patterns of electrical activity in some groups of neurons, that doesn't seem much like saving minds as something distinctive, something we value, something that makes us the feeling, thinking, and rule-following creatures that we are. So why not look outside physicalism? But what options are there if we set aside the physicalist picture? Leaving physicalism behind is to abandon ontological physicalism, the view that bits of matter and their aggregates in space-time exhaust the contents of the world. This means that one would be embracing an ontology that posits entities other than material substances—that is, immaterial minds, or souls, outside physical space, with immaterial, nonphysical properties.[3]

Will a dualist ontology of immaterial minds help us with mental causation and consciousness? That is the question I want to consider here. I will argue that ontological dualism provides us with no help at all, and that in fact it makes things worse. My target will be the interactionist dualism of Descartes. I will be focusing on how mental causation fares within the Cartesian scheme. My conclusion will be: Very badly. As for consciousness, my view is that dualist ontologies offer us no special help; they will only prompt us to reformulate the problem, or perhaps lull us into ignoring it. Brief reflection should convince us that the introduction of immaterial souls as bearers of consciousness will not help to mitigate

2. As things turn out, I will be arguing, in the final chapter, that we have to live with a residual form of epiphenomenalism.

3. Here we will not consider neutral monism and other theories that posit a reality that is neither physical nor mental but of which the physical and the mental are two "aspects" or "manifestations." I believe such theories, in addition to introducing something wholly mysterious and ad hoc, do no better than property dualism, and that in fact for our purposes they will likely turn out to be only variants of property dualism.

William James's sense of perplexity about consciousness,[4] or relieve the emergentists' bafflement over the emergence of consciousness. I will not, however, take up the issue of consciousness in regard to immaterial minds; the difficulties that ontological dualism faces with the problem of causality undermine it so seriously, in my view, as to render the question what further work immaterial souls might do entirely moot.

Cartesian Dualism and Mental Causation

We commonly think that we, as persons, have both a mental and a bodily dimension—or mental aspects and material aspects. Something like this dualism of personhood, I believe, is common lore shared across most cultures and religious traditions, although it is seldom articulated in the form of an explicit set of doctrines as in modern western philosophy and some developed theologies. It is often part of this "folk dualism" that we are able to survive bodily deaths, as souls or spirits, and retain all or most of the mental aspects of ourselves, such as memory, the capacity for thought and volition, and traits of character and personality, long after our bodies have crumbled to dust.

Spirits and souls as conceived in popular lore seem not to be entirely without physical properties, if only vestigially physical ones, and are not what Descartes and other philosophical dualists would call souls or minds—wholly immaterial and nonphysical substances with no physical properties whatever. For example, souls are commonly said to *leave* the body when a person dies and *rise upward* toward heaven, indicating that they are thought to have, and be able to change, locations in physical space. And they can be heard and seen, we are told, by people endowed with special powers and in an especially propitious

4. See chapter 1.

frame of mind. Souls are sometimes pictured as balls of bright light, causing the air to stir as they glide through space and even emitting faint unearthly sounds. But souls and spirits depicted in stories and literature, and in films, are not the immaterial minds of the serious dualist. These latter souls are wholly immaterial and entirely outside physical space.

But can we make sense of the idea that an immaterial soul can be in causal commerce with a material body, and that my immaterial mind can causally influence the physicochemical processes going on in my material brain? Doubts about such a possibility are as old as Descartes's interactionist dualism itself. Conventional wisdom in philosophy of mind has it that its inability to account for mental causation was the downfall of Descartes's mind-body dualism. As has often been noted, his radical dualism of mental and material substances was thought to preclude the possibility of causal transaction between them. Princess Elisabeth of Bohemia achieved philosophical immortality by confronting Descartes with her celebrated challenge to explain "how the mind of a man can determine the bodily spirits [i.e., the fluids in the nerves, muscles, etc.] in producing voluntary actions, being only a thinking substance."[5] According to one commentator, Richard A. Watson, the perceived inconsistency of mind-body causation with the radical duality of minds and bodies was not only a major theoretical flaw in Cartesianism but also the historical cause of its decline.[6]

The reason standardly offered for the supposed incoherence of Cartesian interactionist dualism is that it is difficult to

5. Elisabeth to Descartes, 16 May 1643. This quotation is taken from Daniel Garber, "Understanding Interaction: What Descartes Should Have Told Elisabeth," in *Descartes Embodied* (Cambridge: Cambridge University Press, 2001), p. 172. There is an affecting chapter on Princess Elisabeth in Richard Watson's biography of Descartes, *Cogito, Ergo Sum: The Life of René Descartes* (Boston: David R. Godine, 2002).

6. Richard A. Watson, *The Downfall of Cartesianism 1673–1712* (The Hague, Holland: Martinus Nijhoff, 1966).

"conceive" how two substances with such radically diverse natures, one in spacetime with mass, inertia, and the like and the other lacking wholly in material properties and not even in physical space, could exercise causal influence on each other. Apparently, various principles about causation, such as that cause and effect must show a certain degree of mutual affinity or "essential likeness," that there can be no "greater reality" in an effect than there is in its cause, or that physical causation requires the impact of one body upon another, seem to have played a role. Anthony Kenny, a philosopher well known for his philosophical acuity as well as historical erudition, writes:

> On Descartes' principles it is difficult to see how an unextended thinking substance can cause motion in an extended unthinking substance and how the extended unthinking substance can cause sensations in the unextended thinking substance. The properties of the two kinds of substance seem to place them in such diverse categories that it is impossible for them to interact.[7]

The trouble is that this is all that Kenny has to say about Descartes's difficulties with mind-body causation—and, as far as I know, that is pretty much all we get from Descartes's critics and commentators. But as an argument this is incomplete and unsatisfying. As it stands, it is not much of an argument; rather, it only expresses a vague, inchoate dissatisfaction of the sort that ought to prompt us to look for a real argument. Why is it incoherent to think that there can be causal interaction between things in "diverse categories"? Why is it "impossible" for things with diverse natures to enter into causal relations with one another? What sorts or degrees of diverseness make trouble and why?

It has not been an easy matter to pin down exactly what is wrong with positing causal relations between substances with unlike natures, and explain in concrete terms what it is about

7. Anthony Kenny, *Descartes* (New York: Random House, 1968), pp. 222–23.

the natures of mental and material substance that make them unfit to enter into causal relations with each other. And there have been commentators who have defended Descartes against charges of incoherence like Kenny's. Louis Loeb is one of them.[8] Loeb's defense rests on his claim that Descartes was a proto-Humean about causation—namely that, for Descartes, causality amounted to nothing more than brute regularity, or "constant conjunction," and there can be no a priori metaphysical constraints, such as resemblance or mutual affinity, on what events can be causally joined with what other events. Loeb supports his interpretation with this passage from Descartes:

> There is no reason to be surprised that certain motions of the heart should be naturally connected in this way with certain thoughts, which they in no way resemble. The soul's natural capacity for union with a body brings with it the possibility of an association between thoughts and bodily motions or conditions so that when the same conditions recur in the body they impel the soul to the same thought; and conversely when the same thought recurs, it disposes the body to return to the same conditions.[9]

On Loeb's view, then, the fact that soul and body are of such diverse natures was not, for Descartes, even a presumptive barrier to their entering into the most intimate of causal relations. It seems to me that this reply might be effective as a

8. Louis E. Loeb, *From Descartes to Hume* (Ithaca, NY and London: Cornell University Press, 1981). See pp. 134–49. See also Daniel Garber's "Understanding Interaction: What Descartes Should Have Told Elisabeth," cited in note 5; Eileen O'Neill, "Mind-Body Interaction and Metaphysical Consistency: A Defense of Descartes," 227–45. *Journal of the History of Philosophy* 25 (1987); Marleen Rozemond, *Descartes's Dualism* (Cambridge, MA: Harvard University Press, 1998).

9. *Descartes' Philosophical Letters*, trans. and ed. Anthony Kenny (Oxford: Oxford University Press, 1963), p. 210. I am doubtful as to whether this passage supports Loeb's Humean reading of Descartes, for Descartes is using here causal verbs like "impel" and "dispose" to describe the regularities.

first pass—as a challenge to the critics of Descartes like Kenny to put up a real argument or shut up. Why can't Descartes just say that causation, at least on some fundamental level, is a brute fact based solely on regularities governing events, and that there is compelling evidence, in the form of the countless mind-body correlations familiar from everyday experience, for the reality of mind-body causation? But does it help Descartes to turn him into a proto-Humean "constant conjunctionist" on causation? I don't think it does, and the reason is simple to see and also instructive. It can be seen that Descartes's trouble with mental causation has nothing to do with the bruteness or primitiveness of causation or whether causation is merely a matter of Humean regularity, and that it has everything to do with the supposed nonspatiality of Cartesian minds.

Suppose that two persons, Smith and Jones, are "psychophysically synchronized," as it were, in such a way that each time Smith's mind wills to raise his hand, Jones's mind also wills to raise his (Jones's) hand, and every time they will to raise their hands, their hands rise. There is a constant conjunction between Smith's mind's willing to raise a hand and Smith's hand's rising, and, similarly, between Jones's mind's willing to raise a hand and Jones's hand's going up. If you are a constant conjunctionist about causation, this would suffice for saying that a given instance of Smith's willing to raise a hand is a cause of the subsequent rising of his hand, and similarly in the case of Jones. But there is a problem. For we see that instances of Smith's mind's willing to raise a hand are constantly conjoined not only with his hand's rising but *also with Jones's hand's rising*, and, similarly, instances of Jones's mind's willing to raise a hand are constantly conjoined with Smith's hand's rising. So why is it not the case that Smith's volition causes Jones's hand to go up, and that Jones's volition causes Smith's hand to go up?

It will not do to say that after all Smith wills *his* hand to rise and that's why his willing causes his hand, not Jones's hand, to rise. It isn't clear what this reply can accomplish, but it begs

the question on hand. The reason is that what makes Smith's hand Smith's, not Jones's, that is, what makes Smith's body the body with which Smith's mind is "united," is the fact that there is specially intimate and direct causal commerce between the two. To say that this is the body with which this mind is united is to say that this body is the only material thing that this mind can *directly* affect—that is, without anything else serving as a causal intermediary—and that all changes this mind can cause in other bodies are via changes in this body, changes directly caused by this mind. This is *my* body, and this is *my* arm, because they are things that I can move without moving any other body. I can raise *your* arm only by grabbing it with my hand and pulling it up.[10] And something similar must obtain for body-to-mind causation as well. The "union" of a mind and a body that Descartes speaks of, therefore, presupposes mental causation. Whether or not this is a historically correct reading of Descartes, a causal account of "ownership" seems the most natural option for substance dualists, and I do not know of noncausal alternatives that make any real sense.

I have heard some people say that we could simply take the concept of the mind's "union" with a body as an unexplained and unexplainable primitive,[11] and that it is simply a primitive fact, perhaps divinely ordained,[12] that this mind and this body are merged in a proper union that is a person. I find such an

10. Does this exclude telekinesis? Yes. That probably is the main reason why there is something a priori strange about telekinesis. If telekinesis were a widely spread everyday phenomenon, that might very well undermine the idea that each of us has a distinct body.

11. According to Daniel Garber, something like this was Descartes's view; in addition, Descartes claimed the notion to be intelligible in its own right; see Garber's "Understanding Interaction: What Descartes Should Have Told Elisabeth."

12. For an unabashedly theistic approach of this kind, see John Foster, "A Brief Defense of the Cartesian View," in *Soul, Body, and Survival*, ed. Kevin Corcoran (Ithaca, NY: Cornell University Press, 2001), p. 29.

approach inadequate and unsatisfying. For it concedes that the notion of "union" of a mind and a body, and hence the notion of a person, is unintelligible. For what is it for an immaterial thing wholly outside space to be "united" or "joined" with a material body with a specific location in space? The word "united" merely gives a name to a mystery rather than clarifying it. If God chose to unite my body with my mind, just what is it that he did? I am not asking *why* he chose to unite this particular mind with this particular body, or *why* he decided to engage in such activities as uniting minds and bodies, or *whether* he, or anyone else, could have powers to do things like that. All of that could remain a mystery and I wouldn't complain. What I am asking for is more basic: If God "united" my mind and my body to make a person, there must be a relationship R such that a mind stands in relation R to a body if and only if that mind and that body constitute a unitary person. In uniting my mind and my body, God related the two with R. Unless we know what R is, we do not know what it is that God wrought. The word "union" remains a mere label, and we do not understand what it is that the theistic explanation is attempting to explain when it says that God ordained the "union."

Causation and the "Pairing" Problem

The difficulty we have seen with Loeb's interpretation of Descartes as a Humean in matters of causation, I believe, points to a more fundamental difficulty in the idea that mental substances, outside physical space, can enter into causal relations with objects in physical space. What is perhaps more surprising, the very same difficulty besets the idea that such nonspatial mental substances can enter into any sort of causal relations, whether with material substances or *with other mental substances*.

Let us begin with a simple example of physical causation: two guns, A and B, are simultaneously fired, and this results in

the simultaneous death of two persons, Adam and Bob. What makes it the case that the firing of A caused Adam's death and the firing of B caused Bob's death, and not the other way around? What are the principles that underlie the correct and incorrect *pairings* of cause and effect in a situation like this? We can call this "the causal pairing problem," or "the pairing problem" for short.[13]

Two possible ways for handling this problem come to mind. (1) We can trace a continuous causal chain from the firing of A to Adam's death, and another such chain from the firing of B to Bob's death. (In fact, we can, with a high-speed camera, trace the bullet's trajectory from gun A to Adam and similarly for gun B and Bob.) No causal chain exists from the firing of A to Bob's death, or from the firing of B to Adam's death.
(2) We look for a "pairing relation", *R*, that holds between A's firing and Adam's death and between B's firing and Bob's death, but not between A's firing and Bob's death or B's firing and Adam's death. In this particular case, when the two guns were fired, gun A, not gun B, was located at an appropriate distance from Adam and pointed in his direction, and similarly with gun B and Bob. It is these *spatial relations* (distance, orientation, etc.) that help pair the firing of A with Adam's death and the firing of B with Bob's death. Spatial relations seem to serve as the "pairing relations" in this case, and perhaps for all cases of physical causation involving distinct objects.

The two methods may be related, but let us set aside this question for now.

Turn now to a situation involving nonphysical Cartesian souls as causal agents. There are two souls, A and B, and they perform an identical mental act at time *t*, as a result of which a change occurs in material substance M shortly after *t*. We may

13. The "pairing problem" was first formulated by John Foster in "Psychophysical Causal Relations," *American Philosophical Quarterly* 5 (1968): 64–70. See also Foster's *The Immaterial Self* (London: Routledge, 1991). I earlier discussed this problem in "Causation, Nomic Subsumption, and the Concept of Event," *Journal of Philosophy* 70 (1973): 217–36.

suppose that mental actions of the kind involved generally cause physical changes of the sort that happened in M, and, moreover, that in the present case it is soul A's action, not soul B's, that caused the change in M. Surely, such a possibility must exist. But ask: What relation might serve to pair soul A's action with the change in M, a relation that is absent in the case of soul B's action and the change in M? That is, what could be the pairing relation in this case? Evidently, no spatial relations can be invoked to answer this question, for souls are not in space and are not able to bear spatial relations to material things. Soul A cannot be any "nearer" to material object M, or more propitiously "oriented" in relation to it, than soul B is. Is there anything that can do for souls what space, or a network of spatial relations, does for material things?

But what about mind-to-mind causation? Would this be any easier for Descartes? Consider a purely mental world, a world inhabited only by Cartesian souls; such a world must be possible, since souls are "substances," that is, independent existents. Soul A acts in a certain way and so does soul B at the same time. This is followed by certain changes in two other souls, A* and B*. Suppose that actions of A and B are causes of the changes in A* and B*. But which cause caused which effect? If we want a solution that is analogous to case (2) above for the firings of guns and the deaths, what we need is a pairing relation *R* such that *R* holds, say, for A and A*, and for B and B*, but not for A and B*, or for B and A*. Since these entities are immaterial souls outside physical space, *R* cannot be, or include, a spatial relation, or any other kind of physical property or relation. The radical nonspatiality of mental substances rules out the possibility of invoking spatial relationships to ground cause-effect pairings.

Evidently, then, the pairing relation *R* must be some kind of psychological relation. But what could that be? Could *R* be some kind of intentional relation, such as thinking of, picking out, and referring to? Perhaps, soul A gazes at souls A* and B*,

and singles out A*, and causes a change in it. But how do we understand these relations like gazing at and picking out? What is it for A to pick out A* rather than B*? To pick out some concrete thing outside us, we must be in a certain cognitive relation to it; we must perceive it somehow and be able to single it out from other things near and around it—that is, perceptually identify it. Take perception: What is it for me to perceive this tree, not another tree which is hidden behind it and which is qualitatively indistinguishable from it? The only credible answer we have is the familiar causal account, according to which the tree that I perceive is the one that is causing my perceptual experience as of a tree, and I do not see the hidden tree because it bears no causal relation to my perceptual experience.[14] Ultimately, these intentional relations must be explained on the basis of causal relations (this is not to say that they are wholly analyzable in terms of causality), and this means that we cannot explain what it is for soul A to pick out soul A* rather than B* except by positing some kind of causal relation that holds for A and A* but not for A and B*. If this is right, invoking intentional relations to do causal pairings begs the question: we need causal relations to understand intentionality. Even if intentional relations were free of causal involvements, that would not by itself show that they would suffice as pairing relations. In addition, they must satisfy certain structural requirements; as will become clear as we proceed, they must suffice for the individuation of intentional objects, and it is by no means clear that intentional relations can satisfy these requirements.

We are not necessarily supposing that one single *R* will suffice for all causal relations between two mental substances. But if the physical case is any guide, we seem to be in need of a

14. This of course is the causal theory of perception. See H. P. Grice, "The Causal Theory of Perception," *Proceedings of the Aristotelian Society*, supplementary vol. 35 (1961).

certain kind of "space," not physical space of course, but some kind of a nonphysical coordinate system that gives every mental substance and every event involving a mental substance a *unique location* (at a time), and which yields for each pair of mental entities a determinate relationship defined by their "locations" (analogous to the distance-orientation relation between a pair of spatial objects). Such a system of "mental space" could provide us with a basis for a solution to the pairing problem, and enable us to make sense of causal relations between nonspatial mental entities. But I don't think we have any idea what such a framework might look like—what purely psychological relations might generate such a space-like structure. I don't think we have any idea where to begin.

What about using the notion of causal chain to connect the souls in the right cause-effect relationships? Can there be a causal chain between soul A's action and the change in soul A*, and between soul B's action and the change in soul B*? But do we have any understanding of such purely mental causal chains? What could such chains be like outside physical space? Hume required that a pair of causally connected events that are spatiotemporally separated be connected by a chain of *spatially contiguous* events. It is difficult to imagine what kind of causal chain might be inserted between events involving two mental substances. Presumably we have to place a third soul, C, between soul A and soul A*, such that A's action causes a change in C which in turn causes the change in A*. But what could "between" mean here? What is it for an immaterial and nonspatial thing to be "between" two other immaterial and nonspatial things? In the physical case, it is physical space that gives a sense to betweenness. In the mental case, what could serve the role that space serves in the physical case?

One might say: For soul C to be "between" souls A and A* in a sense relevant to present purposes is for A's action to cause a change in C and for this change to cause a change in A*. That is, betweenness is to be taken as causal betweenness.

This of course is the idea of a causal chain, but it is clear that this idea does not give us an independent handle on the pairing problem. The reason is easy to see: it begs the question. Our original question was: How do we pair soul A's action with a change in soul A*? Now we have two pairing problems instead of one: First, we need to pair soul A's action with a change in a third soul, C, and then we need to pair this change in C with the change in A*. This means that the two methods above, (1) and (2), for cause-effect paring, are not independent, and this for a very simple reason: the very idea of a causal chain makes sense only if an appropriate notion of causation is already in hand, and this requires a prior solution to the pairing problem. It follows that method (2) is the only way to effect cause-effect pairings.

We are, therefore, back with (2)—that is, with the question of what psychological relations might serve the role that spatial relations serve in the case of physical causation. The problem here is independent of the Humean constant conjunction view of causation, and therefore independent of the difficulty we raised for Loeb's defense of Descartes.[15] For suppose that there is a "necessary," counterfactual-sustaining regularity connecting properties *F* and *G* of immaterial mental substances. A mental substance A has *F* at *t*, and an instant later, at *t**, two mental substances, B and C, which share identical intrinsic properties, acquire property *G*. I think we must countenance the following to be a possible situation: A's having *F* at *t* causes B to have *G* at *t**, but it does not cause C to have *G* at *t**. Suppose it is claimed that what distinguishes the two cases is that

15. As can be seen by reflecting on the case discussed earlier of Smith and Jones who are "psychophysically synchronized," the pairing problem arises even at the level of stating Humean constant conjunctions. So, even if we make Descartes into a Humean, as Loeb suggests, this would not help Descartes to escape the pairing problem. More broadly, my arguments do not depend on the use of a heavy-duty concept of causation, as has been suggested by some writers.

the counterfactual "If A had not had *F* at *t*, *B* would not have had *G* at *t**" is true, whereas the counterfactual "If A had not had *F* at *t*, C would not have had *G* at *t**" is false. Well and good. But if that is the case, there must be an intelligible and principled account of why the first counterfactual is true and the second false. I do not believe we could simply assert this as a brute fact for which no explanation is possible or needed and leave it at that. Since B and C are intrinsically alike, the difference in the truth-values of the two counterfactuals must be on account of some relation *R* that A bears to B but not to C. What could this relation be? Cases like this are not outré examples made up for philosophical argument; such situations should be perfectly ordinary ones in a Cartesian world. If so, how would we ascertain causal relationships in such a world?

If these reflections are essentially right, our idea of causation requires that the causally connected items be situated in a space-like framework. It has been widely believed, as we noted, that Cartesian dualism of two substances runs into insurmountable difficulties in explaining the possibility of causal relations across the two domains, mind to body and body to mind—especially, the former. But what our considerations show is that the problem runs deeper: the very same difficulties beset substantival dualism in regard to the possibility of mental-to-mental causation. Under substance dualism, mind-to-mind causation is no more intelligible than mind-to-body causation. Furthermore, the difficulty is rooted deep in the nature of immaterial minds: it is their supposed essential nonspatiality that makes it impossible for them to meet a basic requirement of causality, namely the need for pairing relations. Perhaps, Leibniz was wise to renounce all causal relations between individual substances, or monads—although I have no idea as to his actual reasons for this doctrine. A purely Cartesian world seems like a pretty lonely place, inhabited by immaterial

souls each of which is an island unto itself, totally isolated from all other souls. Even the actual world, if we are immaterial souls, would be a lonely place for us; each of us, as an immaterial entity, would be entirely cut off from anything else, whether physical or nonphysical, in our surroundings. Can you imagine any existence lonelier than an immaterial self?

Causality and Space

The plausible fact that the causal pairing problem for physical causation is solved by invoking spatial relations tells us something important about causation and the physical domain. By locating each and every physical item—object and event—in an all-encompassing coordinate system, the framework of physical space imposes a determinate relation on every pair of items in that domain. Causal structure of the physical domain presupposes this spatial (or more broadly, spacetime) framework. Causal relations must be selective and discriminating, in the sense that there can be two objects with identical intrinsic properties such that a third object causally acts on one but not the other, and, similarly, that there can be two intrinsically indiscernible objects such that one of them, but not the other, causally acts on a third object. We believe that objects with identical intrinsic properties must have the same causal powers or potentials, both active and passive (some would identify the causal powers of an object with the set of its intrinsic properties). However, objects with the same causal powers can differ in the exercise, or manifestation, of their powers, vis-à-vis other objects around them. This calls for a principled way of distinguishing intrinsically indiscernible objects in causal situations, and it is plausible that spatial relations provide us with the principal means for doing this. Prima facie, spatial relations have the right sorts of properties; for example, causal influences generally diminish as distance increases, and barriers

of various sorts can be set up in the right places in space to prevent or impede propagation of causal influences (though not of gravity, we are told). In general, causal relations between physical objects or events appear to depend crucially on their spatiotemporal relations to each other; think about the point of establishing alibis—"I wasn't there," if true, can be sufficient for "I didn't do it." To avoid being burned in a fire, you run away from it as fast as you can—that is, you try to put as much distance as you can between you and the fire. The temporal order alone will not be sufficient as a causal framework; for there can be two or more contemporaneous objects with identical intrinsic properties whose causal behaviors are different. We need a full spacetime framework for this purpose. It was not for nothing that Hume included "contiguity" in space and time, as well as constant conjunction and temporal precedence, among his conditions for causal relations. From our present perspective, Hume's contiguity condition—or the condition that a spatially separated cause-effect be connected by a chain of contiguous cause-effect pairs—can be seen as his solution to the pairing problem. It is also what makes Humean causation an essentially spatial concept. Outside physical space, Humean causation makes no sense. It seems to me that the significance and importance of this condition for Hume's account of causation has not been properly understood or appreciated.

If this is right, it gives us one plausible way of vindicating the critics of Descartes who, as we saw, argued that the radically diverse natures of mental and material substances preclude causal relations between them. It is of the essence of material substances that they have determinate positions in spacetime and that there be a determinate spatiotemporal relationship between each pair of them.[16] Descartes of course talked of extendedness in space as the essence of matter, but we

16. At least on the classic conception of spacetime.

can broadly construe this to include other spatial properties and relations for material substances. Now consider the mental side: As I take it, the Cartesian doctrine has it that it is part of the souls' essential nature that they are wholly outside the spatial order and lack all spatial properties. And it is this essential nonspatiality that makes trouble for their participation in causal relations. As earlier noted, it isn't just mind-to-body causation, but also mind-to-mind causation, that is imperiled by the nonspatiality of immaterial minds.

We have already seen how difficulties arise for mind-to-body and mind-to-mind causation. Unsurprisingly, body-to-mind causation fares no better. The details are similar and can be skipped. But let us note that given that we had trouble envisioning a system of pairing relations for the domain of mental substances, it seems out of the question that we could generate a system that would work across the divide between the mental and material realms. If this is true, not even epiphenomenalism is an option for the serious substance dualist.

I am of course not claiming that these considerations are what motivated the long line of critics of Descartes's interactionism. I am only suggesting a way of fleshing out their worries and showing that there is indeed a concrete basis for these worries. It turns out that, as Kenny and others have said, causal interaction between mind and matter is precluded by their diverse natures, and we have identified the essential diversity that matters, namely the spatiality of bodies and the supposed nonspatiality of minds.

"Affinity" does turn out to make a difference for causation after all. Causality requires pairing relations, and this diversity between minds and bodies does not permit such relations for minds and bodies. What the critics perhaps didn't see was the possibility that the same difficulty bedevils causal relations within the realm of the minds as well. If all this is right, there is no need to appeal to the alleged "mechanical" nature of material causation and the supposed teleological or rational

character of mental causation to show that mind-body causation is problematic. An effective argument can be formulated at a more general and basic level.

Why Not Locate Souls in Space?

These reflections might lead one to wonder whether it would help the cause of dualist causation if immaterial minds were brought into space and given locations in it, not as extended substances like material bodies but as extensionless points. After all, Descartes spoke of the pineal gland as "the seat" of the soul, and it is easy to find passages in his writings that seem to give souls positions in space. And most people, philosophers included, who believe in souls appear to think that our souls are somehow located inside our bodies—my soul in my body and your soul in your body. It seems to me that the thinking here is closely associated with the idea that my soul is in direct causal contact with my body and your soul with your body. The pineal gland is the seat of the soul for Descartes only because it is where unmediated mind-body causal interaction is thought to take place. This confirms my general feeling that mind-body causation generates pressures to bring minds somehow into space, which, for Descartes, is exclusively the realm of matter.

In any case, putting souls into physical space may create more problems than solve them. For one thing, we need a motivated way of locating each soul at a particular point in space. Leibniz said that we locate the soul in a body but we don't think it is at some particular place within it:

> The second [mode of being somewhere] is the *definitive*. In this case, one can "define"—i.e. determine—that the located thing lies within a given space without being able to specify exact points or places which it occupies exclusively. That is how some people have thought that the soul is in the body, because they have not thought it possible to specify an exact point such

> that the soul or something pertaining to it is there and at no other point. Many competent people still take that view.[17]

But does this make any sense? If my soul, as a geometric point, is in my body, it must be either in the top half of my body or its bottom half. If it's in the top half, it must be either in its left or right half, and so on, and we should be able to corner the soul into as small and specific a region of my body as we like. And why should we locate my soul in my body to begin with? Why can't we locate all the souls of this world in one tiny place, say this pencil holder on my desk, like the many thousand angels dancing on the head of a pin?

It would beg the question to locate my soul where my body, or brain, is on the ground that my soul and my body are in direct causal interaction with each other; the reason is that the possibility of such interaction is what is at issue and we are considering the localizability of souls in order to make mind-body causation possible. Second, if locating souls in space is to help with the pairing problem, it must be the case that no more than one soul can occupy a single spatial point; for otherwise spatial relations would not suffice to uniquely identify each soul in relation to other souls in space. This is analogous to the principle of "impenetrability of matter," a principle whose point can be taken to be the claim that space provides us with a criterion of individuation for material things. This principle says that material objects occupying exactly the same spatial region at one time are one and the same—at least from the causal point of view.[18] What we need is a similar principle

17. Leibniz, *New Essays on Human Understanding*, tr. and ed. Peter Remnant and Jonathan Bennett (Cambridge: Cambridge University Press, 1981), bk. 2, ch. 23, sec. 21.

18. This qualification is in consideration of a fairly widely shared view that there are counterexamples ("coincident objects") to the principle that material objects are individuated by spatial coincidence; e.g., a statue and the lump of bronze that "constitutes" it. But then there are philosophers, e.g., Lynne Rudder Baker, who hold—implausibly, in my opinion—that these spatially coincident objects can, and do, have different causal powers.

for souls, that is, a principle of "impenetrability of souls": Two distinct souls cannot occupy exactly the same point in space at the same time. But if souls are subject to spatial exclusion, in addition to the fact that the exercise of their causal powers are constrained by spatial relations, why aren't souls just material objects, albeit of a very special, and strange, kind? Moreover, there is a prior question: Why should we think that a principle of spatial exclusion applies to immaterial souls? To solve the pairing problem for souls by placing them in space we need such a principle, but that is not a reason for thinking that the principle is true. We cannot wish it into truth—we need independent reasons and evidence.

Moreover, if a soul, all of it, is at a geometric point, it is puzzling how it could have enough structure to account for all the marvelous causal work it is supposed to perform and how one might explain the differences between souls in regard to their causal powers. You may say: A soul's causal powers arise from its mental structure, and mental structure doesn't take up space. But what is mental structure? What are its parts and how are the parts configured in a structure? If a soul's mental structure is to account for its distinctive causal powers, then, given the pairing problem and the essentiality of spatial relations for causation, it is unclear how a wholly nonspatial mental structure could account for a soul's causal powers. To go on: If souls exclude each other for spatial occupancy, do they exclude material bodies as well? If not, why not? It may be that one's dualist commitments dictate certain answers to these questions. But that would hardly show they are the "correct" answers. When we think of the myriad questions and puzzles that arise from locating souls in physical space, it is difficult to escape the conclusion that whatever answers might be offered to these questions would likely look ad hoc and fail to convince. Locating souls in space, therefore, is not an option that is going to help the dualist cause, and Descartes was wise to keep them out of it.

Concluding Remarks

I have tried to explore considerations that seem to show that the causal relation indeed exerts a strong pressure toward a degree of homogeneity over its domain, and, moreover, that the kind of homogeneity it requires includes, at a minimum, spatiotemporality, which arguably entails physicality. The more we think about causation, the clearer becomes our realization that the possibility of causation between distinct objects depends on a shared space-like coordinate system in which these objects are located, a scheme that individuates objects by their "locations" in the scheme. Are there such schemes other than physical space? I don't believe we know of any. This alone makes trouble for serious substance dualisms and dualist conceptions of what it is to be a person—unless, like Leibniz, you are prepared to give up causal relations altogether. Malebranche denied causal relations between all finite substances, reserving causal powers exclusively for God, whom he regarded as the only genuine causal agent that exists. It is perhaps not surprising that among the dualists of his time, Descartes was the only major philosopher who chose to include minds as an integral part of the causal structure of the world. In defense of Descartes, we can ask: What would be the point of having souls as immaterial substances if they turn out to have no causal powers, not even powers to be affected by things around them? It is all too easy to excoriate Descartes for his unworkable metaphysics; I believe we should applaud him for showing a healthy respect for commonsense in defending mental causation and in persevering to make sense of our intuitive dualist conception of what it is to be a person. It appears that for many of Descartes's contemporaries, substance dualism was more important than mental causation; Descartes tried to have both and got himself in deep trouble. It seems that at that point in time it didn't occur to most

philosophers to blame the trouble on substance dualism and keep mental causation.

What we must conclude, therefore, is that substance dualism offers little hope for mitigating our difficulties with mental causation. Our considerations show that the very idea of immaterial, nonspatial entities precludes them from entering into causal relations; in fact, I think that the very idea of such objects may well be incoherent and unintelligible. In the preceding two chapters we saw that property dualism, of which nonreductive physicalism is the most influential contemporary form, encounters deep difficulties with mental causation. We have now seen that substance dualism fares no better. Our overall provisional conclusion has to be that no form of dualism, whether substance or property dualism, can serve as a basis for explaining how it is possible for our mentality to be so deeply enmeshed in the causal network of the physical world. For it seems beyond doubt that mentality is part of the causal structure of the world and appears seamlessly integrated into it.[19]

19. This chapter is descended from a paper first presented at a conference on mind-body dualism at University of Notre Dame in March, 1998. See Timothy O'Connor's reply, "Causality, Mind, and Free Will," in *Soul, Body and Survival*, ed. Kevin Corcoran (Ithaca, NY: Cornell University Press, 2001). For further discussion, see Noa Latham, "Substance Physicalism," in *Physicalism and Its Discontents*, ed. Carl Gillett and Barry Loewer (Cambridge: Cambridge University Press, 2001.