REDUCTIONISM'S DEMISE: COLD COMFORT

by Donald H. Wacome

Abstract. Nonreductive physicalism, as opposed to reductionism, enjoys wide popularity by virtue of being regarded as comporting with the traditional image of human beings as free and ontologically unique without the difficulties of mind-body dualism. A consideration of reasons, both good and bad, for which reductionism is rejected suggests instead that the move to nonreductive physicalism does nothing to mitigate the implications of a physicalist account of human nature.

Keywords: freedom; mind-body dualism; nonreductive physicalism; physicalism; reductionism.

The dualistic idea that humans are, in whole or in part, immaterial beings is now widely regarded as outré. To confess it in certain intellectual circles is to reveal a possibly hopeless naivete. Dualism continues to have able and ingenious defenders, but they are fighting a rear-guard action. Physicalism, the view that we are entirely physical beings, has become something close to an academic orthodoxy. The more we learn about the workings of the human brain, the harder it is to insist that the mind is nonphysical. In some quarters, however, physicalism is rarely endorsed without the qualification that, while it is true and good, reductionism is false and bad. In the past it was taken for granted that physicalism and reductionism are more or less synonymous. However, it is now clear that they are distinct: reductionism implies physicalism, but physicalism does not imply reductionism. What makes physicalism palatable to many today is the realization that there can be a *nonreductive physicalism* (Davidson 1980; Post 1987, 159–208). This possibility has been eagerly endorsed. Physicalism is in; reductionism is out. Endorsing reductionism is, for many, as retrograde as endorsing dualism itself.

Donald H. Wacome is Professor of Philosophy, Northwestern College, Orange City, IA 51041; e-mail wacome@nwciowa.edu.

[*Zygon*, vol. 39, no. 2 (June 2004).]
 ISSN 0591-2385

This adverse judgment on reductionism by avowed physicalists is perhaps especially pronounced among persons of faith, for whom a central concern is protecting the traditional human self-conception from the inroads of a scientism intent on denying the uniqueness of human beings. In the course of advocating a physicalist account of the human being, Arthur Peacocke asserts, "Christian believers, affirming as they do the reality, dignity, and value of the human mind, are opposed to reductionism" (1985, 149). More recently, the essays comprising *Whatever Happened to the Soul? Scientific and Theological Portraits of Human Nature* (Brown, Murphy, and Malony 1998) work from the supposition that a physicalist view of human nature that rejects the dualism of body and mind (or soul) comports well with the standpoint of faith so long as it is a nonreductive physicalism. Ian Barbour contends that "without reverting to dualism we can today reject reductionistic materialism and acknowledge human beings as responsible persons" (2002, 129).

What is so bad about reductionism? Does it threaten our reality, dignity, value, or freedom in any way that physicalism as such does not? My aim in this essay is to put forward considerations that undermine the conviction that nonreductive physicalism mitigates the difficulties that the triumph of physicalism poses for our traditional view of ourselves as free and ontologically unique. Physicalism challenges the traditional view of human beings in its nonreductive forms no less than in its reductive forms.

In what follows, I begin with a brief characterization of classical intertheoretic reduction, the strongest version of anything properly called reduction. Then I consider a variety of objections to it and suggest that any merit they have is as objections not to reductive physicalism as such but to physicalism in general. I go on to consider the actual reasons that classical reductionism came to be widely regarded as a failed project and suggest that they offer no clear comfort to those who see it as a threat to the human self-image. Next, I cast doubt on the view that, while reductionism precludes the possibility of free agency, nonreductive physicalism leaves room for it. Finally, I note that the rise of views closely akin to classical intertheoretic reduction threatens to render nonreductive physicalism anachronistic.

Classical Reductionism

The term *reductionism* is used in a variety of ways, but I focus on a core concept. As articulated in the mid-twentieth century heyday of logical empiricism, reductionism was understood as a logical relation between the statements of theoretical languages (Nagel 1979; Hempel 1966, 101–10).[1] A standard example is the reduction of chemistry to physics: the reductionist claimed that the theoretical language of chemistry can be translated into the theoretical language of physics. For instance, the chemist's term

molecule can be translated into physical terms, such as *atom* and *electron*, that refer only to things with which physics is concerned. This translatability is a one-way relation; we cannot in general translate the terms of physics into those of chemistry. The language used by chemists can be replaced with the language used by physicists with no diminution of our power to refer to things, but physicists talk about things for which there is no chemical vocabulary. Some physical things are "invisible" to the language of chemistry; there is no way explicitly to refer to them. A crucial fact here is that reduction is a relation of one set of *types* to another set of types; classical reduction is *type reduction*. If one science is reducible to another, there is a smooth mapping of the types—the concepts, categories, and kinds—of the reduced science onto the concepts, categories, and kinds of the reducing science.

Reducibility was conceived as a logical relation between linguistic items—theoretical vocabularies and the statements framed in them—but type reducibility was taken as having ontological implications. If the languages of chemistry and of physics are completely interchangeable, no more things can be referred to by means of chemists' language than are already referred to by means of physicists' language. If we specify the relation of a particular oxygen atom and two particular hydrogen atoms, particularly the way they share electrons, then we have ipso facto specified a water molecule. We cannot think that there is anything more to a water molecule—this chemical thing—than its physical parts standing in the physical relation in which they stand. After all, a water molecule *is* two hydrogen atoms and an oxygen atom suitably related to one another. So, in a derivative sense, we may speak of the reducibility of one type of thing to another type of thing, of molecules to atoms.

Common Objections to Reductionism

Those who embrace nonreductive physicalism think that it is good and important that reductionism be false. A consideration of widespread objections to reductionism suggests that rejoicing in the possibility of a nonreductive physicalism is ill grounded. Either these objections rely on misunderstandings about what reductionism implies or they apply to *any* sort of physicalism, whether reductive or nonreductive.

Reductionist has a pejorative use in many sectors of contemporary intellectual discourse as equivalent to *simplistic*. Here, a reductionist approach is construed as ignoring things that should not be ignored; it literally reduces the number of things or aspects of things to which we think we need to pay attention. A sociological theory that pays heed to economic facts to the exclusion of other features of a society—its religious and cultural aspects, for instance—might be characterized as reductionistic in this sense. It is true that to think simplistically is a mistake; the question is whether

any particular reductionist theory is in fact too simple, actually leaving out of consideration something that should not be left out. A type-reductionist theory involves a reduction in the number of referring terms and by implication the number of fundamentally different kinds of entity that need to be taken into account. The reduction of chemistry to physics involves the possibility of replacing all of the uniquely chemical vocabulary with that of physics. In this way to reduce is to simplify.

Why should anyone regard the reductionist claim about the in-principle replaceability of one vocabulary with another as simplistic, as ignoring something that should not be ignored? If someone claims that our psychological language can be reduced to a language that talks only about the brain and its doings, what is she ignoring? It's not as though she is saying that in trying to understand the mind we can ignore mental things; presumably she means that we must pay attention to mental things but that they can all be referred to with a strictly physicalist vocabulary. Of course, this implies that those mental things are physical things. Those who believe that the mind is not a physical entity have a right to complain that the reductionist omits something that ought not to be left out—the immaterial mind. However, all physicalist theories of the mind, whether or not they are reductionist, reject the existence of a nonphysical mind and contend that it is the brain, a physical thing, that reasons, chooses, emotes, is self-conscious, and so on. Reductionist theories are not unique in leaving the nonphysical mind out of consideration. This is the essential theoretical simplification that physicalism endorses. Nonreductive physicalism is no less guilty of it than reductionist physicalism is.

A related objection is that reductionist theories of mind deny the reality of the mind. Peacocke appears to be getting at this when he describes reductionism as incompatible with the *reality* of human beings. The contention is that to say that the mind is reducible to something physical is to say that, after all, the mind does not really exist. The reductionist is implicitly characterized as saying that we used to think there were such things as minds, but now we know there aren't any; there are only brains.

The reductionist has a ready response to this objection: at face value it is absurd. Suppose I say that the theoretical language in which Xs are referred to can be reduced to the theoretical language in which only Ys are referred to. The ontological implication of this is that Xs are *identical* to Ys, perhaps to exceedingly complex arrangements of them; any given X is this Y and that Y and that other Y and so on, suitably organized. What cannot reasonably be intended in the reductionist claim is that Xs do not exist *and* that they do not exist because they are identical to Ys, which do exist. That things of a certain sort are the very same things as things that do exist is not a good reason to believe that things of the first sort do not exist. Being the same thing as an existent thing is not the way to achieve nonexistence. That mental things are identical to neuroscientific things is

not a good reason to think that mental things do not exist. What seems to be forgotten here is that reductionism is a theory about the language with which we refer to things and not about the things to which that language refers. To say that the language with which we speak of Xs is in principle replaceable with the language we use to refer to Ys is to make a claim about the referential adequacy of the latter language, not a claim about the nonexistence of the things to which the reduced language refers.

This objection appears to embody a confusion of classical type-reductionism with a theory that is opposed to it, indeed a response to the failure of this sort of reductionism: *eliminativism*. The eliminativist, not the reductionist, contends that the entities referred to by way of the irreducible language should be eliminated from our catalogue of things that really exist.[2]

However, this response is probably too hasty. There are contexts in which the claim that our talk of Xs reduces to talk of Ys is reasonably taken as tantamount to the claim that there are no Xs. Suppose someone believes that Xs *essentially* have such and such a property, some property that Ys essentially lack: he believes that if Xs exist at all, they have this property. He reasonably receives the assertion that nothing has this property as implying that there are no Xs. An analogous case might be that of the traditional theist who regards accounts of language about God as reducible to talk about human ideals and aspirations as tantamount to atheism: she'll say that if, as this reductionist theory claims, there is no supernatural personal being, there is no God. To the one who thinks that the human mind is essentially immaterial, theories on which the mental reduces to the physical are theories in which minds do not exist. On this score the dualist is entitled to object to the reductionist theory, but the objection is properly to physicalism in general, not specifically to its reductive versions. Physicalism, whether reductionist or nonreductionist, denies the existence of the immaterial mind.

Another objection is that the reductionist is saying that Xs—the reduced things—are *nothing but* things of the sort to which the reducing language refers, and this, derisively dismissed as "nothing buttery," is regarded as preposterous (see Peacocke 1993, 40). However, is there a plausible reading of the reductionist claim that Xs are nothing but Ys, which is patently false from a physicalist point of view? It seems unlikely.

Suppose I say of Polly the cat that she is nothing but a cat. I do not mean to say that the predicate *cat* is the only one that legitimately can be applied to her; there are surely any number of other things true of her, even though she is nothing but a cat. Polly is also an organism, a pet, an object that blocks my view of my computer screen, something I often refer to as an example, a thing my friend Jay mistakenly believes possesses a soul, a home for fleas, and so on. I mean that she belongs to a category that excludes her belonging to some other category, whereas someone else

might mistakenly say that, or behave as if, she belongs to that category too. I tell Polly, a creature singularly rich in self-esteem, that she is nothing but a cat, thereby reminding her that she is not a human being. Those who advance reductionist theories of mind are up to something similar, asserting that a human mind is nothing but a material thing, a more or less functioning brain, a collection of suitably interacting elementary particles, not—lest you imagine otherwise—an immaterial substance. Those with dualist inclinations will not appreciate this attempt to put human minds in their ontological place, but there is nothing essentially reductionist in this; all physicalists, reductive or not, regard minds as material.

A related objection is that reductionism implies that the reduced things are *nothing but the sum of their parts*, and that this is false because things are indeed more than the sum of their parts. Three questions arise here: What does it mean to say that a thing is or is not more than the sum of its parts? Does reductionism imply that things are no more than the sum of their parts? and Are the things in question in fact more than the sum of their parts? Consider an ordinary, reasonably complex physical object—an automobile. What would it mean to say that it is more than the sum of its parts? Imagine someone's believing that when all of the parts are put together properly, the result is still not a car, or perhaps not a functioning car; that the car will not work unless another, entirely different sort of part is added—say, a car spirit, an immaterial (or at least very special physical) entity. No one believes this about cars, but for a long time it was almost universally believed about living things. To deny this vitalist view we might say that a biological organism is nothing more than the sum of its parts. To say that is to say something obviously true; there are no vital forces or entelechies; biological life is the result of just putting the physical components together in the right way. This is analogous to what the physicalist says about human beings. When material things of the right kind are assembled in the right way, the result is a human person, mind and all. No further nonphysical thing—no immaterial mind or soul—must be added to do the real work of thinking, sensing, and emoting. Dualists will find this view objectionable, but it is essential to physicalism, whether reductive or nonreductive.

However, the assertion that a thing is no more than the sum of its parts might say something false. Take the car apart, and we wind up with a pile of parts but no car. Some of us who have experimented with the disassembly of complex physical objects have learned the hard way that just collecting all the parts together in one spot is a long way from having the object. If we read *sum* as simply the set of a thing's parts, with no regard to how they are put together, how they relate to one another and interact, it is obvious that most things are more than the sum of their parts. A car is not a set of car parts, even though it has the elements of a set of car parts as its parts. In this sense it would be incorrect to describe a human mind as no

more than the sum of its material parts; a bucket of neurons and glial cells is not a human mind. For the reductionist, things are no more than the sum of their parts, but not in this sense.

Yet another objection is that a reductive account of something *explains it away*. Here too it is difficult to see what could be intended that is at once a reasonable interpretation of the reductionist claim and obviously false, at least from a physicalist point of view. When I hear that something is being explained away, I think of this sort of case: Fred believes that some UFOs are extraterrestrial spacecraft. He experiences a sighting. The Air Force explains it as an atmospheric phenomenon. The UFO has been explained away; the proffered explanation is not what he hoped for or expected.

Whether something has been explained away or simply explained is relative to the sort of explanation wanted or expected in the first place. A vitalist might say of contemporary biology that it has explained away life; he was hoping for a sort of explanation different from what biochemistry and microbiology might supply. Only those committed to an explanation of mental phenomena that posits an immaterial entity are entitled to feel that a reductionist account explains away the mental. Here again the objection is misdirected. It is physicalism in general, not reductionist physicalism per se, that explains away the mental in this sense.

An underlying theme runs through the five preceding objections to reductionism: something has been left out that ought to have been included. It is hard to see what that might be, other than the nonphysical mind, in which dualists believe but which all physicalists reject. I suspect that these commonly voiced objections to reductionism are in reality merely expressions of vestigial discontent with physicalism and have little relevance to the issue of reduction.

A somewhat different type of objection contends that reductionism is committed to untenable views about how the mental can be explained. The core objection here arises in a variety of forms. One version contrasts nonreductive physicalism with the reductionist idea that such higher-level, mental facts can be explained in terms of lower-level, neurobiological facts. Another version contrasts nonreductive physicalism with the reductionist idea that what is true of a complex whole can be exhaustively explained in terms of its parts. Once again, it appears that while these claims are clearly false under some interpretations, the reductionist is committed to them only under other interpretations on which they are plausible from any physicalist perspective, whether or not it is reductionist. Objections of this type seem to miss the significance for issues of explanation of the fact that the mental supervenes upon the physical even if it is not reducible to it.

Physicalism in any form implies that the mental asymmetrically *supervenes* upon the physical. This means that whereas two persons can differ physically without differing mentally, they cannot differ mentally without

there being some physical difference. Whether or not the mental reduces to the physical, the mental facts are fixed by the physical facts; a physical duplicate of this world is necessarily a mental duplicate of it, too.[3] This implies that if we had adequate knowledge of the physical state of the world, including the nature of the supervenience relations that exist between the mental and the physical, we could explain anyone's being in any mental state by pointing out that the world's being in such-and-such a physical state implies that the person is in that supervenient mental state. For the sake of simplicity, suppose that Marvin's mental states supervene locally upon the neurophysiological states of his brain, and ask for an explanation of Marvin's happiness.[4] There is a kind of explanation of his being happy in that his brain is in such-and-such a neurophysiological state, and this implies that he is happy, because happiness supervenes upon that sort of brain state. Similarly, a kind of explanation lies in the fact that earlier Marvin's brain was in a particular neurophysiological state, and this, in conjunction with the relevant laws of nature, implies that he was caused to be in such-and-such a neurophysiological state upon which his being happy supervenes.

What is significant here is that these explanations are possible irrespective of the reducibility of Marvin's mental states to his brain states; they depend only upon the supervenience to which every physicalist is committed. The nonreductive physicalist has no basis for objecting to the mental facts about Marvin—the facts about the whole made up of the trillions of atoms and billions of neurons—being exhaustively explained in this sense. Of course, for finite human minds, such explanations are almost useless. From the reductionist perspective, we make sense of the world scientifically when we detect the higher-level patterns that map onto its lower-level causal regularities. Reductionism's promises about the reducibility of one science to another always related to what is possible *in principle*; no one seriously maintained that we might in practice dispense with the concepts, generalizations, and, if there are any, laws of the higher-level sciences on the grounds of their reducibility to physics.

The claim that reduction implies that the mind can be exhaustively explained in terms of the physical could be interpreted in another way, but when it is taken this way the reductionist does not accept it. It could be construed to mean that the higher-level, mental terms and concepts have no legitimate use in explaining the mind or human behavior, as if reductionism implied that we should reject such explanations of Marvin's happiness as "Marvin wanted to go to the party with Mary, and she accepted his invitation" and replace them with explanations that speak only of Marvin's lower-level, neurophysiological states. The reductionist theory implies that the mental vocabulary is in principle replaceable with the neuroscience vocabulary; this may sound as though the reductionist rejects the former in favor of the latter, but such a view misconstrues the intent of the reduc-

tionist program. To ascertain that the things we refer to when we speak of mental states are the same things that figure in our best scientific theories is not, for reductionism, a reason to discard such concepts. On the contrary, it is a discovery that legitimates their use. From the physicalist point of view, successful reduction serves to root the mind in reality. It is the failure of reduction that casts doubt upon the scientific validity of our mental concepts and, for some, upon the reality of the mind.

Consider two famous examples from the history of science. In the early eighteenth century Georg Stahl posited phlogiston to explain combustion, but this theory ultimately failed to account for the empirical facts and was rejected. Chemistry holds that phlogiston simply does not exist. Statements that refer to it are simply false; they are not translatable into the chemical vocabulary that refers to the various kinds of molecules and their interactions. Phlogiston plays no role in scientific explanation. In contrast, consider a successful reduction, that of heat to the mean kinetic velocity of constituent particles. Statements such as "The water in the beaker has a temperature of 100 degrees" can, in principle, be replaced with complex statements about the molecules in the beaker and their aggregate motion. Heat, unlike phlogiston, has a place in the world portrayed by modern science and a role in scientific explanation precisely because it is reducible to the lower-level language of physics and chemistry. Chemistry textbooks, undeterred by heat's reducibility, are replete with assertions about heat and its effects. In general, the reducibility of a pretheoretical concept to some scientifically acceptable theoretical vocabulary underwrites its explanatory validity.[5] The successful reduction of mind to brain would show why it is entirely reasonable to go on explaining human behavior by appealing to beliefs, desires, emotions, and all of the other familiar mental entities. As Paul and Patricia Churchland put it, a successful reduction "vindicates" the reduced theory (1998, 73).

Reductionism's Demise: Cold Comfort

It might seem perverse to dwell on the failings of these popular objections to classical intertheoretic reductionism in light of the fact that it is widely regarded as a refuted theory. However, as will become clear, those who look to its demise as making physicalism safe for the traditional human self-image make mistakes that parallel those embodied in the objections just cited. The move from reductionist to nonreductionist physicalism offers no comfort to those concerned with reductionism's impact on our self-image.

Even as classical reductionism was being perspicaciously formulated, its end was in sight. For a generation it has been clear that as a general thesis about relations among the sciences classical reductionism is untenable. There might be some intertheoretical reductions; possibly chemistry does

reduce to physics, but in general it appears not to work.[6] There is, for instance, a good case to be made for biology's not reducing to chemistry. The kind *gene*, a crucial biological type, arguably cannot be defined in chemical terms. All the genes we know of happen to be segments of DNA molecules, yet *gene* is a *functional* concept. A gene is anything, no matter what it happens to be made of, that acts like a gene—that is, that plays a certain information-theoretic role in cellular reproduction. There are any number of ways to make something that plays that role; genes could be made of Tinkertoys® or Play-Doh® or trained cats. The kind *gene* is multiply realizable: there are many different ways to be a gene, so many that there is no hope of spelling them out in the language of physics.

Reductionism, in its failure, separates from physicalism and reveals the possibility of a nonreductive physicalism. Consider the question of whether such folk-psychological concepts as *belief* and *choice* and *desire* embodied in our everyday language about the mind reduce to neuroscience, and, thus, whether the mind is ultimately reducible to the things of which physics speaks. It may have seemed that, while a successful reduction implies physicalism, the irreducibility of mind to physical reality would imply dualism; but the relation is not symmetrical. Successful reduction counts decisively in favor of physicalism, but reductive failure does not give us a good reason to think that the irreducible things are immaterial. To discover irreducibility is to discover nothing particularly significant about the ontology of the mind.

To appreciate why, consider an ordinary material object, say, a pizza, and consider the kind-term *pizza* that we use to refer to pizzas. Two things seem indisputable. One, we cannot reduce pizza talk to physics talk; pizzas cannot be defined in the vocabulary of physics. In that parsimonious language we cannot list the individually necessary and jointly sufficient conditions that something must satisfy to be a pizza. Here again the issue is multiple realizability: there are too many ways to be a pizza. If there is a pizza essence, it cannot be expressed by means of the conceptual resources of physics. Pizzas are among the world's irreducible realities.

Yet, pizzas indisputably are physical things. Pizza irreducibility inspires no one to embrace pizza dualism; there is nothing more to a particular pizza than the physical stuff of which it is made configured in the way it is configured. The fact that we cannot define what it is to be a pizza by speaking only of quarks and electrons does not imply that any given pizza is anything more than a large number of quarks and electrons suitably arranged. We are all nonreductive physicalists when it comes to pizzas and to most of the objects of everyday life. To say that the mind is irreducible to physics, or to some higher-level science such as neurophysiology, is to say something about it that is true about most, maybe all, macroscopic material things. The irreducibility of the mental to the physical seems about as ontologically interesting as the irreducibility of pizzas to the physical

and is not likely to point to solutions to the problems our being material poses for traditional conceptions of what it means to be human. Nonreductive physicalism offers no obvious route to human uniqueness.

Nonreductive Physicalism And Freedom

Perhaps nonreductive physicalism's strongest attraction lies in the hope that it makes possible an account of free agency consistent with our being material beings. Typically we explain the choices that human beings make by appealing to their reasons for those choices, that is, to their beliefs and desires. We explain, for instance, Karen's choosing to go to the zoo by pointing out that she wanted to see an aardvark and believed that the best place to see one was at the zoo. The crucial question is whether our reasons *cause* our choices and, if they do, whether this is compatible with those choices' being free. From the incompatibilist perspective we can be free only if our reasons do not cause our choices.

On plausible assumptions, the type irreducibility of the mental to the physical implies that there are no causal laws that refer to mental states as such (Davidson 1980; Fodor 1981). The laws of nature are about the causal relations between physical types, not the irreducible types of our psychological vocabulary. No causal law dictates that if Karen wants to see an aardvark and believes that the best place to see one is at the zoo she will choose to go to the zoo. This might inspire someone to think that in escaping classical, intertheoretic reduction we have made the world safe for incompatibilist free choice. However, this move runs afoul of the fact that, even if reduction fails and there are no type identities, there still might be *token* identities. Even if mental kinds (types) are not identical to physical kinds (types), it might still be the case that any particular mental thing is numerically identical to—the very same individual thing as—some particular physical thing.

There is no classical reduction of talk about pizzas to talk about the quarks and electrons that they are made of; the kind *pizza* is not identical to any physical kind. Nonetheless, it is highly plausible that any given pizza is *token* identical to a particular arrangement of quarks and electrons existing at a particular place and time. Likewise, we can reject the type identity of the mental and the physical and still say that any given mental event is token identical to some physical event. So long as the mental is in this sense identical to the physical, the mind remains decisively implicated in the cause-and-effect matrix of the natural world. Dispensing with classical reduction and type identity does nothing to change that. Once again there seems to be nothing special about human minds. There are no laws of nature about such irreducible items as pizzas or pencils, yet these are surely governed by such laws. The term *pencil* appears in no law of nature, but this does not exempt my pencil from falling to the floor when I drop it

as prescribed by the law of gravity. So long as the mental is token identical to the physical, the mind appears to be completely embedded in the causal structure of nature just as everything else in the physical world is.[7]

We should note in passing that there is good reason to worry that if mental events are not at least token identical to the physical events that cause bodily movements, the mind lacks causal efficacy: it would be an illusion that our conscious choices cause our behavior. Indeed, some think that this epiphenomenalist consequence is avoidable only if some sort of type reduction and type identity obtains (Kim 1998). The cost of ensuring that choices are not identical to physical events, and thus not the effects of other physical events, might be their being causally impotent.[8] The conclusion that our choices have no physical effects goes far beyond securing the view that they have no physical causes.

Nancey Murphy, a prominent advocate of nonreductive physicalism, seeks to safeguard freedom by denying the identity of the mental and the physical without denying physicalism itself. She asks, "If human behavior is entirely reducible to chemistry, and chemistry to physics, then is it not the case that the laws of physics ultimately determine everything we do and that human free will is an illusion?" (Murphy 1997, 13) She attempts to demonstrate the nonidentity of the mental and the physical by pointing out that mental states not only possess the property of being multiply realizable but are also "multiply constituted"—that is, the same type of brain state can be the supervenience base for different types of mental state, depending upon the circumstances in which the mental state is realized. This suffices to show that mental types are not identical to neurophysiological types and, thus, that no classical reduction succeeds (Murphy and Ellis 1996, 32–37; Murphy 1998, 128–43).

However, multiple constitutibility is consistent with the token identity of particular mental events with particular brain events. Consider the example Murphy offers, in which two experimental subjects receive a small electric shock to the back. Subject A has been led to expect to have ice applied to his back (Murphy 1998, 137). Because of this "mental set," he experiences the shock as a sensation of cold. Subject B has been led to expect something hot to be applied to his back; his mental set leads him to experience the shock as the sensation of being burned. We are to assume that the application of the shock to the back produces the same kind of neurophysiological event in both subjects. Thus we have:

- A's sensation, *w*: a particular event, a token of the mental type *C*, "sensation of cold"
- A's brain event, *x*: a particular event, a token of *N*, a neurophysiological type
- B's sensation, *y*: a particular event, a token of the mental type *H*, "sensation of heat"

- B's brain event, *z*: a particular event, a token of *N*, a neurophysiological type

This shows that *N* is not identical to *C* and that *N* is not identical to *H*. It does not show that *w* is not numerically identical to *x* or that *y* is not numerically identical to *z*; token identity of mental event and brain event might obtain in the absence of any identity of mental types and neurophysiological types. Two identical brain events, *x* in A and *z* in B, could be sensations of two different types.[9]

This is important for the issue of freedom, because whatever causally determines a particular brain event also determines any mental event numerically identical to it. The mental events we call choices might well be type irreducible to some type of brain event, but if any given choice is token identical to a particular brain event, whatever causes that brain event is its cause. The nonidentity of the mental and the physical for which Murphy contends appears to be of no help in the attempt to secure incompatibilist freedom. It leaves the human mind fully embedded in the cause-and-effect physical world.

The token identity of mental and physical particulars is widely accepted among physicalists. However, not everyone accepts it. It might be that, although mental events supervene upon the brain, it is no easier to identify mental particulars with neurophysiological particulars than it is to identify mental types with neurophysiological types (Haugeland 1998). Suppose that Mike is thinking about bananas and is doing so because of what is going on in his brain; his banana thought supervenes upon his brain. He cannot stop thinking about bananas and start thinking about beer without there being appropriate changes in his brain, and his brain's being in this state is sufficient for his thinking about bananas. This supervenience relation may hold even though his banana thought cannot be identified with any particular brain event, even if his banana thought turns out to depend upon brain activity so widely distributed that there is no plausible candidate for the neurophysiological event or state that is this mental event. This possibility acquires some credibility from a connectionist, or parallel distributed processing, model of the mind. On this account the mind, consisting of a network of many interconnected processors (neurons), represents bananas not because it contains an explicit banana symbol but by virtue of the overall pattern of weighted connections. Little less than the current state of the whole neural network can be identified with a specific representational state (Bechtel and Abrahamsen 1991, 1–20,147–75).

However, abandoning the token identity of the mental and the physical appears to offer no more to those intent on procuring freedom for material persons than does giving up on type identities. Even if there are no useful identifications of supervenient mental particulars with anything less than the state of the whole brain, this will not alter the fact that the brain and all

its states are fully subject to natural law. Supervenience shorn of both type identities and token identities offers nothing to support the conclusion that mental events somehow escape the net of physical causation and, thus, nothing to encourage the view that nonreductive physicalism can give us the kind of freedom incompatibilism demands. Physicalism needs help from neither the type identity nor the token identity of the mental and the physical to pose problems for our traditional conception of ourselves as free agents. Those problems arise from physicalism itself.

Reductionism Redux?

Reductionism is perhaps not so bad after all; at least it is no worse than nonreductive physicalism, except for the rather large impediment of being generally regarded as false. Even this view might be passing. There are accounts of the relation between the mind and the brain that are reductionist in spirit yet apparently immune to the difficulties that encumbered classical reductionism.

We should first recognize that multiple realizability implies the nonidentity of mental and physical types only on the assumption that the types specified in the reducing language must be relatively simple—not wildly complex, messily gerrymandered disjunctions that would be of no use in scientific explanation. If this requirement were to be relaxed, a physicalist might have no reason to deny the reducibility of the mental to the physical; any mental property or entity or event would be identical to some physical property or entity or event. The classical notion of intertheoretic reduction originated in the philosophy of science, so it is not surprising that this requirement has generally been seen as decisive in its failure. However, philosophers whose concerns are more generally metaphysical might see no reason to regard reductionism as a failed project (see, for example, Lewis 1994, 412–15).[10]

Also, it is not clear that the collapse of classical reductionism spells doom for the type reducibility of the mental to the physical as such. Multiple realizability might defeat the view that, for example, the very general mental type *pain* is identical to a particular type of human brain state; other kinds of organism might experience pain even though their brains are constructed on quite different principles than ours. However, type reductions might be possible if restricted to a single species. Possibly the mental type *human pain* is, after all, identical to some type of human brain state (Bechtel and Mundale 1999; Kim 1998, 93–97).

Further, we should keep in mind that *functionalism*, the main successor to the mind-brain identity theory and to the type reductionism involved in it, is a theory that, while principally motivated by the multiple physical realizably of mental types, does not typically reject type reducibility as such. Instead, it replaces the project of reducing mental types to neurophysi-

ological types with the project of reducing mental types to functional types specifiable as causal roles. Functionalism is the theory of mind that underlies the computational approaches to the mind that, despite their difficulties, dominate cognitive science. Suppose that mental types reduce to computational types and that computational types, being realizable in any number of quite different physical systems, cannot be reduced to neurophysiology: this hardly gives aid in the quest for human ontological uniqueness or comfort to those who seek to extricate the human mind from the causal structure of the physical world. At face value, it leaves us as deeply embedded in this world governed by causal law as the machine running the irreducible, multiply realizable word-processing software that I am using to write this.

Finally, it is important to recall that the conception of reduction that has for a generation seemed generally implausible treated theories as essentially linguistic in form and reducibility as a logical relation between sets of propositions. Failures of reduction were failures of translatability. Contemporary "new wave" accounts of intertheoretic reduction dispense with the linguistic element and construe reduction as a structural isomorphism between mathematical models (Hooker 1981; Bickle 1998). Arguably, these are not subject to the difficulties that overwhelmed linguiform reductionism. Should mental to physical reduction in this sense appear to be on the way to success, confident espousals of nonreductive physicalism would sound increasingly hollow.

Conclusion

The disrepute into which reductionist versions of physicalism have fallen can obscure the fact that physicalism itself is far from cost free, so far as our traditional self-understanding goes (Bielfeldt 2001). As far as our conception of what we are is concerned, there may be much that is bad about reductionism, but what's bad about it is just what's bad about physicalism. There may well be no plausible account of human nature that is genuinely physicalist but avoids the conclusion that we are as embedded in the physical world as all other material things are. Whatever value, dignity, and freedom we possess are the value, freedom, and dignity possible for physical things, irrespective of the success or failure of any reductionist project. Nonreductive physicalism is not a secure place to stand. We must either go back to defending dualism or get on with the task of ascertaining how much of our traditional self-image survives the discovery that we are material beings.

NOTES

A version of this essay was presented at the Upper Midwest Regional Meeting of the American Academy of Religion in St. Paul, Minnesota, in April 2002.

1. What I characterize here as classical, intertheoretic reduction is roughly a conjunction of what Barbour (2002, 20) characterizes as *ontological* and *epistemological* reduction. The uses of *reductionism* and its cognates vary widely, especially in the literature on religion and science. One can find oneself appreciating the quip that the term's principal use is to denote whatever aspects of materialism a particular materialist dislikes.
2. A parallel question worth asking, but one that I will not go into here, is whether the demise of eliminativism would do anything to preserve the traditional human self-image once physicalism is accepted.
3. Different versions of supervenience depend on precisely what kind of necessity is involved here (Kim 1993).
4. The more realistic supposition might be that his mental states supervene upon this brain in its environment and with its particular history.
5. It is a further question when something's irreducibility disqualifies it from having an explanatory role.
6. Even the paradigmatic cases of allegedly successful reduction face significant difficulties on close inspection (see, for example, Sklar 1992, 106–17; Stewart and Cohen 1997, 33–61).
7. Some nonreductive physicalists believe that the fact that the mental supervenes upon but does not reduce to the physical world opens the possibility of a downward causation in which mental events act upon the physical world without themselves being determined by physical events. This is regarded as refuting the view, labeled *causal reductionism*, that all of the higher-level characteristics of a complex whole are determined by the lower-level events on which they supervene (Murphy 1998, 129). Although it is highly plausible that in some sense there is such a thing as downward causation—the state of my brain a few moments from now will in part be an effect of my thinking the thoughts I am thinking right now—it is difficult to imagine how this could be of use in securing incompatibilist free choice. On physicalist assumptions mental events are events in the physical world and thus have physical causes; they are neither uncaused nor caused by events occurring in something nonphysical. For the case that there is little metaphysical mileage to be derived from the idea of downward causation see Bielfeldt 1999.
8. Oddly, some defenders of nonreductive physicalism assert that it is the *identity* of the mental and the physical that would render the mental epiphenomenal (Polkinghorne 1996, 18; Murphy 1998, 132).
9. The case is analogous to that of a computer in which two exactly similar events in the hardware might play quite different computational roles, depending on what software the machine is running.
10. While successful reduction can have significant ontological implications, suggesting that there are fewer fundamental kinds of thing than appearances might lead us to expect, more broadly it is important to keep in mind that the facts about reducibility and irreducibility are primarily facts not about the world but about the thought and language with which we refer to the world. Once we make the move to physicalism, the ontological status of the mind is essentially decided; further facts about the reducibility or irreducibility of the language with which we refer to it cannot change that.

REFERENCES

Barbour, Ian. 2002. *Nature, Human Nature, and God*. Minneapolis: Fortress.

Bechtel, William, and Jennifer Mundale. 1999. "Multiple Realizability Revisited: Linking Cognitive and Neural States." *Philosophy of Science* 66 (June): 175–207.

Bechtel, William, and Adele Abrahamsen. 1991. *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*. Oxford: Basil Blackwell.

Bickle, John. 1998. *Psychoneural Reduction: The New Wave*. Cambridge: MIT Press.

Bielfeldt, Dennis. 1999. "Can Downward Causation Make the Mental Matter?" *CTNS Bulletin* 19 (Fall): 11–21.

———. 2001. "Can Western Monotheism Avoid Substance Dualism?" *Zygon: Journal of Religion and Science* 36 (March): 153–77.

Brown, Warren S., Nancey Murphy, and H. Newton Malony, eds. 1998. *Whatever Happened to the Soul? Scientific and Theological Portraits of Human Nature*. Minneapolis: Fortress.

Churchland, Paul M., and Patricia S. Churchland. 1998. "Intertheoretic Reduction: A Neuroscientist's Field Guide." In *On the Contrary: Critical Essays, 1987–1997*, 65–79. Cambridge: MIT Press.

Davidson, Donald. 1980. "Mental Events." In *Essays on Action and Events*, 207–25. Oxford: Clarendon.

Fodor, Jerry. 1981. "Special Sciences." In *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, 127–45. Cambridge: MIT Press.

Haugeland, John. 1998. "Weak Supervenience." In *Having Thought: Essays in the Metaphysics of Mind*, 89–107. Cambridge: Harvard Univ. Press.

Hempel, Carl G. 1966. *Philosophy of Natural Science*. Englewood Cliffs, N.J.: Prentice-Hall.

Hooker, C. A. 1981. "Towards A General Theory of Reduction. Part I: Historical and Scientific Setting. Part II: Identity in Reduction. Part III: Cross-Categorical Reduction." *Dialogue* 20:38–59, 201–36, 496–529.

Kim, Jaegwon. 1993. "Concepts of Supervenience." In *Supervenience and Mind: Selected Philosophical Essays*, 53–78. Cambridge: Cambridge Univ. Press.

———. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge: MIT Press.

Lewis, David. 1994. "Lewis, David: Reduction of Mind." In *A Companion to the Philosophy of Mind*, ed. Samuel Guttenplan, 412–31. Oxford: Basil Blackwell.

Murphy, Nancey. 1997. *Reconciling Theology and Science: A Radical Reformation Perspective*. Kitchener, Ont.: Pandora.

———. 1998. "Nonreductive Physicalism: Philosophical Issues." In *Whatever Happened to the Soul? Scientific and Theological Portraits of Human Nature*, ed. Warren S. Brown, Nancey Murphy, and H. Newton Malony, 127–48. Minneapolis: Fortress.

Murphy, Nancey, and George F. R. Ellis. 1996. *On the Moral Nature of the Universe: Theology, Cosmology, and Ethics*. Minneapolis: Fortress.

Nagel, Ernest. 1979. *The Structure of Science: Problems in the Logic of Scientific Explanation*. Indianapolis: Hackett.

Peacocke, Arthur. 1985. "A Christian 'Materialism'?" In *How We Know*, ed. Michael Shafto, 146–68. New York: Harper and Row.

———. 1993. *Theology for a Scientific Age: Being and Becoming—Natural, Divine, and Human*. Minneapolis: Fortress.

Polkinghorne, John. 1996. *The Faith of a Physicist*. Minneapolis: Fortress.

Post, John F. 1987. *The Faces of Existence: An Essay in Nonreductive Metaphysics*. Ithaca, N.Y.: Cornell Univ. Press.

Sklar, Lawrence. 1992. *Philosophy of Physics*. Boulder, Colo.: Westview.

Stewart, Ian, and Jack Cohen. 1997. *Figments of Reality: The Evolution of the Curious Mind*. Cambridge: Cambridge Univ. Press.