



Materialism

Andrew Melnyk*

Materialism is nearly universally assumed by cognitive scientists. Intuitively, materialism says that a person's mental states are nothing over and above his or her material states, while dualism denies this. Philosophers have introduced concepts (e.g., realization and supervenience) to assist in formulating the theses of materialism and dualism with more precision, and distinguished among importantly different versions of each view (e.g., eliminative materialism, substance dualism, and emergentism). They have also clarified the logic of arguments that use empirical findings to support materialism. Finally, they have devised various objections to materialism, objections that therefore serve also as arguments for dualism. These objections typically center around two features of mental states that materialism has had trouble in accommodating. The first feature is *intentionality*, the property of representing, or being about, objects, properties, and states of affairs external to the mental states. The second feature is *phenomenal consciousness*, the property possessed by many mental states of there being something it is like for the subject of the mental state to be in that mental state. © 2012 John Wiley & Sons, Ltd.

How to cite this article:

WIREs Cogn Sci 2012, 3:281–292. doi: 10.1002/wcs.1174

INTRODUCTION

Nearly all cognitive scientists take it for granted that there is no more to a thinking, feeling, and reasoning person than his or her body, where a person's 'body' is understood, of course, to include his or her brain. In philosophy of mind, however, this assumption of materialism has been brought into the open as the object of explicit investigation. Philosophers of mind have sought to formulate the assumption with more precision, to map the positions in logical space that deny the assumption, and to construct and evaluate both arguments that provide reason to endorse the assumption and arguments that provide reason to reject it. This article aims to introduce non-philosophers to some of the main results of these efforts. But rather than survey the voluminous and often highly esoteric philosophical literature, it aims to provide a sophisticated yet comprehensible framework for actually thinking about the various questions that materialism raises. While the results of philosophical thought on the topic of materialism are likely to have minimal significance

for the daily work of cognitive scientists in the immediate future, they should be of interest to all cognitive scientists who wish to understand the place of the mind in the material world at the highest level of abstraction as well as at the level of concrete detail.

THE MIND–BODY PROBLEM

The central problem in traditional philosophy of mind is the so-called *mind–body problem*. To a first approximation, the problem is to answer the following question:

What is the exact nature of the relationship between, on the one hand, someone's mental (or psychological) condition at a given time and, on the other hand, the person's neurophysiological condition at that time?

For example, what is the exact nature of the relationship between your having a visual experience with a certain introspectible character right now and your being in whatever exact neurophysiological state you are in fact in right now? What is the exact nature of the relationship between your understanding right now what you are reading and your being in whatever exact neurophysiological state you are in fact in right now?

*Correspondence to: melnyka@missouri.edu

Department of Philosophy, University of Missouri, Columbia, MO, USA

Certain features of how the question above has just been formulated are inessential. The formulation above speaks of mental and neurophysiological *conditions* and *states*, but a parallel question could equally well be asked about mental and neurophysiological *processes*. The formulation above speaks of people, i.e., *human* bearers of mental states, but the question applies equally to whatever non-human (e.g., chimpanzee) bearers of mental states there might be. The formulation above implies that the only kind of material state that might be relevant to determining a person's mental state at a given time is a neurophysiological state of that person at the same time; but there are philosophical views concerning the representational content of mental states which imply that what mental state someone is in at a given time is determined not merely by the simultaneous neurophysiological state that the person is in but also by the physical environment that the person is in, and indeed by the history of his or her interaction with his or her physical environment.

The formulation above assumes that mental states exist, i.e., that people really are in mental states. But not all philosophers of mind accept this assumption. Some of those who do not—the so-called eliminative materialists—hold that mental states are merely the posits of a folk theory of behavior that is radically false, and hence that mental states no more exist than do the medieval humors.¹ Others of them—the so-called instrumentalists—hold that the human practice of attributing mental states to one another is merely a fiction that serves us very well in predicting human behavior.² In this article, however, it will be assumed that mental states exist.

To state the question that expresses the mind–body problem, it is not necessary to define the words ‘mental’ and ‘material’ (or ‘physical’). If pressed to say what we mean by ‘mental states’, we can reply simply by listing examples of mental states: thinking that lunch is at noon, wanting to visit Paris, feeling pain, feeling jealousy, seeming to see a patch of blue, and so forth. If pressed to say what we mean by ‘material states’, we can reply by giving examples of uncontroversially material states: neurophysiological states, neurophysiologically implemented computational states, physicochemical states, and so on. It might be valuable for some purposes to know necessary and sufficient conditions for a state to count as mental, but we do not need to know them to understand, or to address, the mind–body problem.

The correct answer (or answers) to the question that expresses the mind–body problem cannot be discovered by performing some simple experiment.

Nor can it just be ‘read off’ textbooks or journals in psychology or cognitive neuroscience or whatever. For very cautious language is used in such places to describe the relation between mental states and neurophysiological states. A recent journal paper is representative: it speaks of ‘brain areas *involved in* pain processing’ and ‘the neural *basis of* pain processing’ (italics added).³ Philosophers of mind have tried to lessen the imprecision of such expressions as those italicized.

MATERIALISM AND DUALISM

In trying to solve the mind–body problem, philosophers of mind have carefully distinguished various possible views concerning how mental states might be related to material states. These views may be classified as *dualist* or as *materialist*. Intuitively, dualist views imply that mental states are something over and above material states, whereas materialist views imply that mental states are nothing over and above material states.

Materialism

The contributions of many philosophers of mind have led to the articulation of two ways in which a mental state might be so related to a material state that the former could be said to be nothing over and above the latter.

The Type-Identity Way

Let us call the first way the *type-identity* way. Your being in pain at time *t* is related to your being in material (e.g., neurophysiological) state *M* at *t* if, and only if, the following condition is met:

- pain = *M*.

Philosophers often call a kind of state a *type* of state, or *state type*. In this jargon, the above condition can be expressed by saying that pain is the very same state type as *M*. The sameness in question is not exact similarity, a relation that typically holds between *two* entities; it is the relation of *being the very same thing as*, a relation that can only hold between one thing and itself (the one thing might have two *names*, of course). If pain is the very same state type as *M*, then it is not even logically possible for a creature to be in pain without being in *M*, or to be in *M* without being in pain. Hence your being in *M* at *t* logically entails your being in pain at *t*, and conversely. After all, if pain = *M*, there is only *one* state type at issue, even though we may refer to it by using either of

the two expressions, 'pain' and 'M'. An important corollary is that someone who claims that $\text{pain} = M$ is not denying the existence of pain. On the contrary, someone who claims that $\text{pain} = M$ (and who accepts that M exists) *cannot* deny that pain exists, for a state that does not exist can hardly be one and the same as a state that does exist. As noted above, there are philosophers who deny the existence of such things as pains and other mental states; but those who assert the identity of mental-state types with material-state types are not among them. To be a materialist it is not necessary to be an eliminative materialist.

Such claims as that $\text{pain} = M$ are not intended as, and should not be interpreted as, claims whose truth or falsity can be evaluated by *a priori* reflection on the meanings of the words 'pain' and 'M'. They should be understood on the model of such scientific identity claims as that being alcohol = being $\text{C}_2\text{H}_5\text{OH}$, having consumption = being infected with *Mycobacterium tuberculosis*, being at such-and-such temperature = having so-and-so mean molecular kinetic energy, being a flame = being a certain mixture of reacting gases and solids that emits light of various wavelengths, and so on. These claims were hard-won empirical discoveries, not the results of *a priori* reflection.

Mental-state types may be contrasted with what philosophers call mental-state *tokens*. Mental-state types are repeatable, i.e., indefinitely many people can be in the same mental-state type (e.g., pain) at indefinitely many times. But mental-state tokens, by definition, are unique, one-off occurrences. Examples of mental-state tokens are your understanding what you are reading right now, my experiencing a dull backache for 1 h from noon yesterday, and Smith's thinking all her adult life that no planets lie outside the solar system. A mental-state token is always a token of some mental-state type. Indeed, a given mental-state token is always a token of more than one mental-state type. For example, my experiencing a dull backache for 1 h from noon yesterday was a token of the more specific type, *experiencing a dull backache*, as well as of the more general type, *being in pain*. Even highly specific mental-state types are repeatable; hence they should not be confused with mental-state tokens.

There is a systematic connection between claims that mental-state *types* are identical to material-state types, and claims that mental-state *tokens* are identical to material-state tokens. For example, if $\text{pain} = M$, then, for every token of the mental-state type, *pain*, there must be a token of the material-state type, M , with which the token of pain is identical. Note, however, that, even if, for every token of the mental-state type, *pain*, there is a token of some or other

material-state type with which the token of pain is identical, these tokens of some or other material-state type may not all be of the *same* material-state type—which is to say that *pain* may not be identical with *any* material-state type.

The Realization Way

Let us call the second way in which a mental-state token might be so related to a material-state token that the former could be said to be nothing over and above the latter the *realization* way. Your being in pain at time t is related to your being in material (e.g., neurophysiological) state M at t in this way if, and only if, the following three conditions are met:

- $\text{pain} \neq M$,
- $\text{pain} =$ the state of being in one of the state types that play causal role R , and
- M is one of the state types that play causal role R .

Such claims as that $\text{pain} =$ the state of being in one of the state types that play causal role R , just like such claims as that $\text{pain} = M$, should be understood as ones whose truth or falsity cannot in general be evaluated by *a priori* reflection on the meanings of the words; they must be, or be entailed by, empirical hypotheses.

Some philosophers would wish to add a fourth condition to the three stated above:

- your being in pain at time $t =$ your being in M at t .

But this seems unnecessary. Given that you are in M at t , then, given the third condition, you must be in one of the state types that play role R . But in that case, given the second condition, you must be in pain at t . Your being in pain at t is therefore nothing over and above your being in M at t . The relation that holds in this case between your being in M at t and your being in pain at time t is what some philosophers call *realization*; when the three conditions are met, your being in M at t can be said to *realize* your being in pain at time t . Furthermore, because the second condition implies that there is more than one state type that plays causal role R , different pain state tokens on different occasions might be realized, respectively, by material-state tokens that belong to different material-state types. If different pain state tokens on different occasions are in fact so realized, then pain can be said to be *multiply realized*.

It might be wondered why the fourth condition noted above could not simply *replace* the second and

third conditions. But very few, if any, philosophers today take this possibility seriously.

One reason is that no one has been able to explain how it even could be the case that your being in pain at t = your being in M at t given that (1) pain $\neq M$ and (2) the second and third conditions above do not hold either; it seems incoherent.

For expository ease only, the second condition has been formulated more restrictively than it ought to be. It says that pain is the state of being in one of the state types that play causal role R . This claim comprises two subclaims. The first subclaim is that pain is a so-called *higher-order* state type, i.e., a state type to be in which just is to be in one of the state types that meet some specific condition. The second subclaim is that this specific condition is the playing of a particular causal role; for a state type to play a particular causal role is for tokens of the state type to be caused by so-and-so causes and in turn to cause such-and-such effects. When philosophers speak of a *functional* state type, they very often mean a higher-order state type for which the defining specific condition is the playing of a certain causal role; similarly, when philosophers speak of *functionalism*, they very often mean the view that mental-state types are one and the same as certain functional state types. It is this second subclaim, however, that makes the formulation of the second condition unduly restrictive. It is not necessary to characterize a higher-order state type as the state of being in one of the state types that *plays a certain causal role*. In principle, a higher-order state type could be characterized in various other ways, e.g., as the state of being in one of the state types that *have a certain biological function*, or that *stand in certain non-causal relations to other states*, or that *constitute a certain stage in the implementation of a certain programme*. It is the identification of pain with a *higher-order* state type of some sort or other that is doing all the metaphysical work here, i.e., explaining how your being in pain at t can be nothing over and above your being in M at t , even though pain $\neq M$.⁴

We have now seen two ways in which a mental state might be so related to a material state that the former could be said to be nothing over and above the latter. Some readers, however, may be aware of philosophical claims to the effect that mental phenomena *supervene* on material phenomena. Intuitively, the idea is that mental states supervene on material states if, and only if, any two entities that are exactly the same with regard to the material states they are in must also be exactly the same with regard to the mental states they are in—i.e., any two entities that differ mentally in any way must also differ materially in some way.⁵ But few philosophers

of mind today think that claims of supervenience provide a third way in which a mental state might be so related to a material state that the former could be said to be nothing over and above the latter. A claim of supervenience tells us nothing about the relationship between one's mental state at a given time and one's material state at the same time—except that the former somehow necessitates the latter. Now if this necessitation relation is taken to hold *because* one's mental state at t is related to one's material state at t in one of the two ways already noted, then the claim of supervenience has contributed nothing to solving the mind–body problem. But if, on the other hand, this necessitation is taken to be a *sui generis* fundamental relation holding between two state tokens of two distinct state types, it is then at the very least dubious that one's mental state at t is nothing over and above one's material state at t ; a *sui generis* fundamental relation of necessitation could hold, it seems, between one's material state at t and as spooky an immaterial state of oneself as could be imagined.⁶

A Formulation of Materialism

It is now possible to formulate the thesis of materialism. First we stipulate what it is for a particular mental-state type, e.g., pain, to be *materialistically acceptable*:

Pain is materialistically acceptable if, and only if, either of the following two conditions is met:

- (1) there is some material-state type, m , such that pain = m and
- (2) there is some higher-order state type, h , such that pain = h , and every state token of h is realized by some or other state token of some or other material-state type.

Then we can define materialism as the view that *all mental-state types are materialistically acceptable*.⁴ It will also be useful to define a state type as *immaterial* if, and only if, it is neither material nor materialistically acceptable.

Dualism

The alternatives to materialism are, of course, the various forms of dualism. Traditionally, philosophers have distinguished between two main forms of dualism, substance dualism and property dualism. *Substance dualism* is the view that mental-state types are one and the same as certain immaterial-state types that certain immaterial substances are in. ('Substance' is the traditional philosophical term

for an entity that persists through time, capable of gaining and losing properties.) *Property dualism* is the view that mental-state types are one and the same as certain immaterial state types that certain *material* substances, presumably brains, are in; on this view, then, brains possess immaterial as well as material properties. However, we should acknowledge the logical possibility of at least one more form of dualism, which we could call *functionalist dualism*: the view that mental-state types are one and the same as certain *higher-order* state types, (some of) the state tokens of which are realized, however, by state tokens of some or other *immaterial* state type.

Now we can imagine a version of property dualism on which there is no system at all to the way in which material entities are in mental states. On such a view, it might be that some humans are in mental states while other humans—who might even be their molecule-for-molecule duplicates—are not; it might even be that rocks are in mental states. Were the imagined dualist view true, mental properties would be more or less randomly distributed over material entities. Actual defenders of property dualism, however, reject this imagined view, instead holding a form of property dualism usually called *emergentism*. On emergentism, only material systems that exhibit a certain degree and kind of (material) complexity can be in mental states, and which mental states such systems are in depends on the exact material states that they are in. More precisely, the mental states of a system are said to *emerge* from material states of the system in accordance with certain fundamental *laws of emergence*; according to such laws, whenever a material system of such-and-such a material kind enters so-and-so material state, it simultaneously enters such-and-such (immaterial) mental state. It is important to see why such laws must be fundamental. Because emergentists are property dualists, they must deny that emergent mental-state types are identical with material-state types (including those from which they emerge), or that they are identical with higher-order state types whose tokens are realized by tokens of material-state types. But in that case laws of emergence cannot be explained in terms of underlying laws of physics, since physical laws only govern *material* states. So laws of emergence must in that sense be fundamental.

The term ‘emergent’, both in philosophy and outside it, is used in many different senses, and is often used in ways that do *not* entail that an emergent phenomenon is materialistically unacceptable; e.g., phenomena are often called emergent when they look like they must have been created by an intelligent agent

but in fact arose from rule-governed local interactions among the elements of a complex system. These usages are, of course, merely different, not wrong; the important thing is to distinguish emergentist views that understand ‘emergent’ in such ways from the dualist view stated above, which is the usual bearer of the name ‘emergentism’ in contemporary philosophy of mind.

EVIDENCE FOR MATERIALISM

The progress of science over the past hundred or so years has yielded evidence of various kinds that materialism of some sort or other is true. Materialism entails that whatever causes the firing of our motor neurons, and hence our outward (including linguistic) behavior, is a material object; since this behavior is remarkably complex, materialism therefore predicts that this material object must itself be of corresponding complexity. And this prediction has been borne out, since obviously science has discovered that the proximate cause of the firing of our motor neurons is the brain, which is indeed an organ of immense complexity. There is nothing inevitable about our having discovered this. We can imagine having discovered instead that the skull contained only cerebrospinal fluid, or that the brain was only as complex as, say, the heart, in which case we would surely have concluded that something *immaterial* must cause our motor neurons to fire, since nothing material is a remotely plausible candidate to do so. Again, materialism entails that humans only possess such mental capacities as a purely material, i.e., materialistically acceptable, object could possess, and hence it predicts that no human is in fact capable of certain logically possible forms of telepathy or telekinesis. And science has systematically failed to show that humans possess these mental capacities. Had we discovered instead that humans are capable of, say, feats of learning from others that no purely material system could manage, we would have had to conclude that humans are not purely material. Finally, materialism predicts that purely material systems of sufficient complexity could exhibit highly elaborate and unpredictable behavior; this prediction looks unlikely to be borne out if one’s paradigm material object is a rock. However, the use of scientific discoveries to build computer-controlled artifacts with remarkable behavioral capacities does in fact bear it out. Imagine that technology had yielded no machines whose behavioral capacities exceeded that of, say, a 1950s washing machine; this would have been important evidence against materialism.

The Causal Argument

Philosophical discussion of the state of the evidence for materialism, however, has mainly focused on just two lines of empirically based reasoning. The first is sometimes called the *causal argument*, but it is more aptly called the *argument from causal closure*.⁷ The argument can be formulated in several non-equivalent ways, but here is one version.⁸ It begins with the commonsense assumption that our intentions, or decisions, or beliefs and desires often *cause* those bodily motions (e.g., muscle contractions) that make up our behavior. On this assumption, many mental states have neurophysiological events among their effects. Presumably these mental states are not causally sufficient all by themselves for their neurophysiological effects; but they *are* causally sufficient for their effects *in the circumstances in which they actually occur*, circumstances that would be insufficient alone. However, the great success of neurophysiology to date in explaining neurophysiological phenomena is strong evidence that the neurophysiological effects of mental events already have sufficient neurophysiological causes; that is, the realm of the neurophysiological is *causally closed*, meaning that to causally explain these neurophysiological effects it is in principle never *required* to leave this realm by invoking a non-neurophysiological cause. It follows that the neurophysiological effects of mental states have both *mental* causes that are sufficient (in the circumstances) for these effects and *neurophysiological* causes that are sufficient for these effects.

On the dualist view that mental states are immaterial, however, the mental cause of a given neurophysiological effect is neither a token of a mental-state type identical with some material-state type nor a token of a higher-order state type that is realized by some or other material-state token; the mental cause is therefore *entirely distinct* from any neurophysiological (or other material) cause of its effect. On the dualist view, then, the neurophysiological effects of mental states have two sufficient causes, one mental and the other neurophysiological, each entirely distinct from the other; they are therefore *overdetermined*.

However, this conclusion is one that we should avoid if we can, for two reasons. First, if we construe mental causes as entirely distinct from neurophysiological causes, then we sacrifice theoretical parsimony without gaining any ability to causally explain anything that otherwise could not be causally explained. Second, consider the unfortunate victim of a firing squad whose death is causally overdetermined when he is struck by two bullets each

of which was sufficient in the circumstances to kill him; in this case, the convergence of two separate causes to produce the same effect can be explained—the firing squad was deliberately organized, after all. But if a mental cause and a neurophysiological cause converge to produce the very same neurophysiological effect, we must treat this convergence as a coincidence and, indeed, if such convergence occurs routinely, as a coincidence that occurs over and over again.

In fact, of course, we *can* avoid the conclusion that the neurophysiological effects of mental states are overdetermined—by rejecting the dualist view and adopting instead the materialist view that the mental causes of neurophysiological effects are materialistically acceptable, i.e., that they are *either* tokens of mental-state types identical with material-state types *or* tokens of higher-order state types that are realized by material-state tokens. For if the mental causes of neurophysiological effects are nothing over and above simultaneous neurophysiological causes, then mental causes operate entirely in virtue of the neurophysiological events that underlie them and there is no overdetermination. In light of the causal closure of the neurophysiological, then, materialism is more likely than dualism.

Responses

Various responses to the argument from causal closure are open to dualists. One response is to deny the commonsense assumption that our mental states do cause the bodily motions that make up our behavior; this requires commitment to the *epiphenomenalist* view that, although mental states are effects, they are never causes (at least of physical effects). Another response is to insist that it is still an open empirical question whether the neurophysiological effects of mental events really do have sufficient neurophysiological causes. A third response is to concede that materialism is preferable to dualism in light of the causal closure of the neurophysiological, but to claim that, because of various philosophical considerations, dualism is preferable to materialism, all things considered. A fourth response—philosophically the most interesting—is to charge that materialism too has difficulty accommodating the fact that mental states cause neurophysiological effects, since, it is argued, if materialism is true, then the mentality of mental causes is never relevant as such to their effects.⁹ Materialists, of course, contest this charge.

The Argument from Neurophysiological Dependence

Let us turn now to the second of the two main lines of empirically based reasoning in favor of

materialism that have been discussed recently in the philosophical literature.¹⁰ The crucial premise in this reasoning is that in recent decades innumerable correlations between mental states and activities, on the one hand, and neurophysiological states and activities, on the other, have been discovered by the use of such techniques as fMRI. In one version of the argument, what these empirical findings make plausible is the claim that the mental is *dependent* on the neurophysiological in the specific sense that, for any person you like, and for any type of mental state that person might be in or mental process that person might undergo, in order for that person to be in that type of mental state or undergo that mental process, there is neurophysiological activity of some distinctive kind that has to be going on—simultaneously—in that person's brain. Two points about this claim deserve emphasis. First, the scope of the claim is not limited to some restricted class of mental states or activities that are relatively primitive or concrete, say, sensory states; it applies to *all* types of mental state or activity, including the most abstract and sophisticated, such as doing mental arithmetic or thinking that God exists.¹¹ Second, the claim is not that all types of mental state or activity depend on some *single* kind of simultaneous neurophysiological activity, in the way in which all types of mental state or activity depend on the functioning of the subject's circulatory system; the claim is rather that each type of state or activity requires a *different* kind of simultaneous neurophysiological activity (there can be overlap, of course).

How do materialism and dualism compare when it comes to explaining the empirically discovered dependence of the mental on the neurophysiological? The answer is that, while both views can be construed so as to explain it, materialism does so more parsimoniously. Consider materialism first. If all mental-state types are materialistically acceptable, then the dependence of the mental on the neurophysiological is precisely what one would expect to find. For if a mental-state type is identical with a neurophysiological-state type, then obviously there cannot be a state token of the former type without a state token of the latter type. And if—the only other possibility—a mental-state type is a higher-order state type whose tokens are all realized by neurophysiological-state tokens, then there cannot be a state token of the mental-state type without a state token of some or other material-state type. Either way, the mental requires the neurophysiological.

Now consider dualism. It is true that the empirically discovered dependence of the mental on the neurophysiological does refute certain forms of

dualism, e.g., those holding that, while the brain is needed for the mind to receive sensory inputs and to produce motor outputs, the mental activity that occurs in between requires nothing whatever from the brain. But dualism could take a different form: it could treat mental-state types as immaterial, and then posit hitherto unrecognized *fundamental laws of nature* that connect each (immaterial) mental-state type to the neurophysiological-state type on which it depends. This form of dualism could explain the dependence of the mental on the neurophysiological.

So both materialism and dualism can be made to entail the dependence of the mental on the neurophysiological. But—and this is a crucial methodological point—it does not follow that materialism and dualism are both equally well supported by it. Two hypotheses may both entail a certain body of data and yet be differentially supported by it.¹² A spectacular example is provided by Philip Gosse's notorious 1857 hypothesis that God created the world rather recently, but made it appear to be millions of years old. This hypothesis entails the very same geological data that Gosse's geologist contemporaries accounted for by hypothesizing that the Earth is very old and that unimaginably slow geological processes operated to produce its modern appearance; but no one finds these two hypotheses to be equally credible in light of the geological data. Similar cases are pervasive. Take any hypothesis, scientific or everyday, that we accept on the strength of a given body of evidence; then we can usually dream up a rival hypothesis that entails the same body of evidence but that we reject as less credible.

What feature might make one hypothesis more credible than a second in light of evidence that both hypotheses entail? In some cases the feature might be a better fit with background knowledge or greater explanatory power, but the feature most clearly relevant to the case at hand is *parsimony* (or economy). Other things being equal, the more parsimonious of two hypotheses is more probable than the less parsimonious. And a hypothesis is more parsimonious than a second hypothesis to the extent that it is committed to the existence of fewer basic kinds of entity and to the holding of fewer basic laws. Now materialism is obviously more parsimonious than dualism. First, dualism is committed not only to all the neurophysiological-state types to which materialism is committed but also to—what dualism considers quite distinct—all the very many mental-state types that there are. Second, dualism is committed not only to all the standard physicochemical laws to which materialism is committed but also to countless fundamental psychophysical laws that it posits to

explain the empirically discovered dependence of the mental on the neurophysiological. The conclusion is that, in light of the dependence of the mental on the neurophysiological, materialism is more probable than dualism—even than epiphenomenalist dualism, we might add, since the causal efficacy of the mental has not been assumed.

Responses

The argument from the dependence of the mental on the neurophysiological has been challenged on the grounds that, while inference to the best explanatory hypothesis is legitimate in general, the claim that mental-state types are identical with neurophysiological (or higher-order) state types is not a proper explanatory hypothesis.^{13,14} A second possible response, as above, is to concede that materialism is preferable to dualism in light of the dependence of the mental on the neurophysiological, but to claim that, because of various philosophical considerations, dualism is preferable to materialism, all things considered.

OBJECTIONS TO MATERIALISM

Objections to materialism have typically focused on two particular features of mental states: *original intentionality* and *phenomenal consciousness*.

Original Intentionality

Such mental states as thinking that Cleveland is north of Chicago, wanting to eat ice cream, or fearing that the car has no gas exhibit what philosophers call intentionality. While intentions to do so-and-so (e.g., to buy milk) certainly exhibit intentionality, intentionality is *not* the property of being intended, and, as the examples illustrate, mental states that are not intentions can still exhibit intentionality. The intentionality of a mental state is its being *about*—or *directed toward*—something, typically something distinct from itself, e.g., an object, stuff, or state of affairs. But the intentionality of a mental state seems not to be just a relation between the mental state and that toward which the mental state is directed. For one can think something that is false (Cleveland is not north of Chicago) and want what does not exist (e.g., to meet Santa Claus). How can one stand in a relation to states of affairs or objects that do not exist? Moreover, in the well known story, Lois Lane believes that Superman is exciting while not believing that Clark Kent is exciting. How could these be different beliefs, given that Superman = Clark Kent, and that any material relation Lane stands

in to Superman she also stands in to Clark Kent? Such questions are why the intentionality of mental states is *prima facie* problematic for materialism.

Now sentences of natural languages also exhibit intentionality, but they are clearly material. So what is the problem? The answer is that, according to a popular view, the intentionality of sentences is somehow derived from the intentionality of mental states; crudely, the idea is that the sentence, 'Grass is green', means that grass is green because there is a population among whom there exists a convention by which a speaker utters 'Grass is green' when he or she wants to get a hearer to think that he or she thinks that grass is green. But obviously the intentionality of mental states cannot be derived in turn from that of anything else; the intentionality of mental states must be *original*, i.e., underived.

Materialist philosophers have proposed reductive accounts of the nature of original intentionality.^{15–17} Such accounts aim to state individually necessary and jointly sufficient conditions for a state to exhibit some sort of intentionality, e.g., to represent that grass is green or to refer to dogs, where these conditions could in principle be met by a purely material system. There are major differences between different accounts, but they all appeal to some combination of the following kinds of ingredients: causal relations, dependence in accordance with laws of nature, covariation, isomorphism, biologically normal conditions, and functions. One crucial task that such accounts must accomplish is to allow for the occurrence of misrepresentation, as when, e.g., someone *mistakenly* thinks that so-and-so. For example, Jerry Fodor aims to state a sufficient condition for a mental symbol S to refer to cows.¹⁵ Roughly, he proposes that tokens of S refer to cows if (1) the presence of cows produces tokens of S and (2) if anything that is not a cow (e.g., an elk on a dark night) produces tokens of S, then, e.g., elks would not have produced tokens of S unless cows had done so, whereas cows would still have done so even if elks had not. From Fodor's proposal it follows that, even if some elk in poor light causes me to think mistakenly that there is a cow over there, the mental symbol activated in my thought can still refer to cows. The important teleological account of intentionality developed by Ruth Millikan accounts for misrepresentation in a quite different way, but it is too complex to be explained here.¹⁷

All reductive accounts of the nature of original intentionality proposed thus far have been subjected to extensive criticism, often taking the form of counter examples; and no account has won more than a small

circle of adherents. Some dualists are apt to think that all these accounts fail, and that the best explanation of this history of failure is that no such account can be given, even in principle, because original intentionality is an immaterial property of the states that have it. Materialists, on the other hand, even when they agree that all these accounts fail, are more optimistic, holding that some account appealing to the sorts of ingredients mentioned in extant accounts will one day turn out to be correct; the project of trying to give such accounts, they note, is only a few decades old.

Phenomenal Consciousness

Sensations (or experiences) such as being in pain or having a red afterimage, as well as those accompanying veridical perceptions, are said by philosophers to be phenomenally conscious. A mental state is phenomenally conscious if, and only if, there is something it is like for the subject of the state to be in that state, something of which we are aware in introspection. Thus, there is something it is like for someone in pain to be in pain, and it differs from what it is like for someone with an itch to have an itch, from what it is like for someone smelling gasoline to smell gasoline, and so forth. The commonest reason why philosophers reject materialism is that there are powerful arguments for the conclusion that what it is like for someone in a phenomenally conscious mental state to be in that mental state is not a materialistically acceptable property. Space permits treatment of only one such argument, and the most influential and widely discussed of them is the so-called Knowledge Argument.¹⁸

The Knowledge Argument

According to it, we can imagine a super scientist of the future, Mary, who has never actually experienced pain herself, but who has come to know everything that a completed multidisciplinary science of pain has to say about pain. She therefore knows all the physics and chemistry of the various kinds of stimuli that cause pain, all the neurophysiology of nociceptors, the somatosensory cortex, the anterior cingulate cortex, and the rest, and all the causal and representational properties of the pertinent neuronal events. But Mary does not know *everything* about pain. For imagine that, once her studies are over, Mary stubs her toe and thus experiences pain for the first time in her life. In doing so, she *learns* something: she learns what it is like to have a pain in the toe. Therefore, since before she felt her first pain she did not know what it is like to have a pain in the toe, but did know everything that a

completed multidisciplinary science of pain has to say about pains in the toe, what it is like to have a pain in the toe must be a property that even a completed multidisciplinary science of pain fails to mention. Such a property cannot in that case be a material property, and so materialism is false.

The Knowledge Argument has generated a huge critical literature. But before we turn to the most influential objection to the argument let us pause to consider the relevance of the argument to an important aspect of the phenomenology of phenomenally conscious mental states. When one attends to one's own phenomenally conscious mental states, it is natural to think of oneself as aware of a private realm (of colors and shapes, say)—private in the sense that not only is no other person actually aware of it but no other person even could be aware of it, no matter what he or she was able to learn about one's behavior, one's environment, the current condition of one's brain, or the functional roles played by particular brain states that one is in. That we all think of our own phenomenally conscious mental states in this way presumably explains why we never feel that we can be *completely* certain how the phenomenally conscious mental states of other people compare with our own, when, say, they and we are both looking at exactly the same ripe tomato. It seems quite impossible, however, for any material (or materialistically acceptable) state to be such that only one person, even in principle, can have cognitive access to it; it seems essential to what is material (and materialistically acceptable) that it be in principle cognitively accessible to more than one person. So if it really is true that, in attending to one's own phenomenally conscious mental states, one is aware of a private realm, it follows that phenomenally conscious mental states are not material (or materialistically acceptable). Materialists must therefore deny that we are aware of a private realm when we attend to our own phenomenally conscious mental states, and must hold also that we are in error when we think that we are.

But what accounts for our thinking of ourselves as aware of a private realm to which no one else can possibly have access? A plausible answer is that, when we attend to our own phenomenally conscious mental states, we have exactly the same intuition that leads us to judge, in the Mary case, that Mary still learns something after she feels pain for the first time, despite her already knowing everything that a completed multidisciplinary science of pain has to say about pain. For what Mary knows about pain before she feels it for the first time is *precisely* the maximum that *other* people could know about one's own pain;

it is the most that can be known from the third-person (as opposed to first-person) point of view. The intuition is simply that what can be known about us, even ideally, from the third-person point of view still omits something that we can each know about ourselves through introspection. The corollary is that this something must concern a reality to which each of us alone, and no one else, has cognitive access, even in principle. Now if this explanation of why we think of ourselves as aware of a private realm when we attend to our own phenomenally conscious mental states is correct, then a successful materialist diagnosis of where the Knowledge Argument goes wrong promises also to reveal the error in our thinking of ourselves as aware of a private realm to which no one else can possibly have access.

Response to the Knowledge Argument

The most influential materialist objection to the Knowledge Argument has been to challenge the validity of the inference *from* the claim that when Mary stubs her toe she learns what it is like to have a pain in the toe *to* the conclusion that what it is like to have a pain in the toe is a property that is not mentioned in a completed multidisciplinary science of pain.^{19,20} The inference is invalid, it is said, because Mary, in learning what it is like to have a pain in the toe, might in fact be learning, of her pain, that it has a certain *material* property that a completed multidisciplinary science of pain *already* speaks of, albeit in the technical language of the neurosciences, and that Mary therefore already knew of. In that case, Mary, in knowing what it is like to have a pain in the toe, would simply be mentally representing this material property in a format different from the one she used to represent it before she first stubbed her toe.

These different representational formats might go along with differences in computational roles sufficiently large that it would be appropriate to speak of Mary's genuinely having *learned* something after she stubs her toe, even though there was no new property of which she learned the existence. Here is an analogy. Imagine that, because of a blow to the head, I suffer terrible amnesia and forget my name. But I read in the newspaper that tomorrow, for reasons that do not matter, one Andrew Melnyk will be publicly flogged. 'Bad news for Melnyk', I think to myself, but soon return to my quest to find out who I am. Later, however, I discover that *I* am AM, and then, of course, I realize, to my horror, that tomorrow *I* will be publicly flogged! The point is that, when I realize that tomorrow *I* will be publicly flogged, I really do learn something, even though I already knew that AM

will be publicly flogged, and my being flogged is the very same occurrence as AM's being flogged, so that I do not learn of some occurrence of which I was previously unaware.

A possible reply to this objection is that, even if the story about different formats that represent the very same property is true, it still does not explain why Mary did not already know what it is like to have a pain in the toe before she stubbed her toe. Why could she not have used empirical methods to convince herself before stubbing her toe that what it is like to have a pain in the toe just is so-and-so material property, so that she could then use her complete knowledge of material properties to infer before stubbing her toe what having a pain in the toe is like?

At this point, materialist objectors to the Knowledge Argument need to say more about the different way in which, after she has stubbed her toe, Mary mentally represents that material property of her pain in the toe that, according to materialism, just is what it is like to have a pain in the toe. They do so by positing the existence of what are termed *phenomenal concepts*.¹⁹ On this view, concepts in general are mental representations of objects, stuffs, or properties, and are the constituents of mental representations of complete states of affairs (e.g., a thing's having a certain property). Phenomenal concepts in particular are supposed to be the concepts that subjects use to represent their own phenomenally conscious mental states while they are actually in those states, and attending to them introspectively; the subjects might then report what they are thinking, albeit imperfectly, with such sentences as 'My current experience is like that' or 'It's like that with me now', but they need not.

Now materialist objectors to the Knowledge Argument hypothesize that phenomenal concepts have a special feature that enables them to explain why Mary could not have used empirical methods to convince herself before stubbing her toe that what it is like to have a pain in the toe just is so-and-so material property. The special feature is that a subject cannot possess (and hence use) a phenomenal concept of a given phenomenally conscious mental state unless the subject is actually in that state at the time, or has at least been in that state in the past. So, to return to Mary: if, after she has stubbed her toe, Mary thinks about her pain by using a phenomenal concept of pain, then, since this is the first pain that she has ever experienced, Mary *cannot* previously have possessed (and hence used) this phenomenal concept, and so she cannot previously have come to think (by empirical or any other means) that what it is like to have a

pain in the toe just is so-and-so material property. However, this appeal to phenomenal concepts is highly controversial. Whether there can be a plausible account, consistent with materialism, of how a concept could have the special feature just described, and whether phenomenal concepts even exist, is the topic of active research.²¹

There is other work that an appeal to phenomenal concepts might be able to do. For example, *philosophical zombies* are molecule-for-molecule physical duplicates of actual people who are in phenomenally conscious mental states, but there is nothing it is like to be these duplicates. David Chalmers has developed an elaborate argument against materialism which assumes that philosophical zombies are conceivable, and which infers from this that, in a sense he spells out, they are really possible; such a possibility is incompatible with materialism.²² Some philosophers, however, think that the hypothesis of phenomenal concepts can be used to show that philosophical zombies might be conceivable even though they are not possible in any sense at all—which would refute Chalmers' argument.²³ Further discussion of

this matter is unfortunately beyond the scope of this article.

CONCLUSION

While consensus on whether materialism is true is as far away in philosophy as it has ever been, progress has nonetheless been achieved. The commitments of materialism and dualism have been clarified. The ways in which empirical findings from the sciences of the mind are actually or potentially able to support—or undermine—materialism are better understood than in the past. And the materialist project of trying to account for the apparent peculiarities of phenomenally conscious mental states by hypothesizing special features of the first-person ways in which we mentally represent these mental states opens up exciting possibilities for interdisciplinary collaboration in the future. For it may be that these hypotheses can be developed to the point of empirical testability, in which case cognitive science might be able to contribute to solving the mind–body problem in a hitherto unimagined way.

REFERENCES

- Churchland P. Eliminative materialism and the propositional attitudes. *J Philos* 1981, 78:67–90.
- Dennett D. *The Intentional Stance*. Cambridge, MA: MIT Press; 1987.
- Schweinhart P, Bushnell MC. Pain imaging in health and disease—how far have we come? *J Clin Invest* 2010, 120:3788–3797. doi:10.1172/JCI43498.
- Melnyk A. *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge, UK: Cambridge University Press; 2003.
- Kim J. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge, UK: Cambridge University Press; 1993.
- Horgan T. From supervenience to superdupervenience: meeting the demands of a material world. *Mind* 1993, 102:555–586.
- Peacocke C. *Holistic Explanation: Action, Space, Interpretation*. New York: Oxford University Press; 1979.
- Melnyk A. Some evidence for physicalism. In: Walter S, Heckmann HD, eds. *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. Exeter, UK: Imprint Academic; 2003, 155–172.
- Kim J. *Mind in a Physical World*. Cambridge, MA: MIT Press; 1998.
- Hill CS. *Sensations: A Defense of Type Materialism*. Cambridge, UK: Cambridge University Press; 1991.
- Harris S, Kaplan JT, Curiel A, Bookheimer SY, Iacoboni M, Cohen MS. The neural correlates of religious and nonreligious belief. *PLoS ONE* 2009, 4:e7272. doi:10.1371/journal.pone.0007272.
- Laudan L. Damn the consequences! *Proc Address Am Philos Assoc* 1995, 69:27–34.
- Kim J. *Physicalism, Or Something Near Enough*. Princeton, NJ: Princeton University Press; 2005.
- McLaughlin B. Consciousness, type physicalism, and inference to the best explanation. *Nous Suppl Philos Issues* 2010, 20:266–304.
- Fodor J. *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press; 1990.
- Dretske F. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press; 1988.
- Millikan RG. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press; 1984.
- Jackson F. What Mary didn't know. *J Philos* 1986, 83:291–295.
- Horgan T. Jackson on physical information and qualia. *Philos Q* 1984, 34:147–151.
- Loar B. Phenomenal states. *Nous Suppl Philos Perspect* 1990, 4:81–108.
- Alter T, Walter S, eds. *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness*

- and Physicalism*. New York: Oxford University Press; 2007.
22. Chalmers D. *The Character of Consciousness*. New York: Oxford University Press; 2010.
23. Balog K. Acquaintance and the mind-body problem. In: Hill CS, Gozzano S, eds. *New Perspectives on Type Identity: The Mental and the Physical*. Cambridge, UK: Cambridge University Press; 2012.

FURTHER READING

- Kim J. *Philosophy of Mind*. 3rd ed. Boulder, CO: Westview Press; 2010.
- Papineau D. *Thinking about Consciousness*. New York: Oxford University Press; 2004.
- Tye M. *Consciousness Revisited: Materialism without Phenomenal Concepts*. Cambridge, MA: MIT Press; 2009.