

THE MIND-BODY PROBLEM

In Defense of Mind–Body Dualism

BRIE GERTLER

Brie Gertler works primarily on philosophical problems in epistemology and the philosophy of mind. She teaches at the University of Virginia.

How quaint, the idea that our minds somehow float free of the cold, hard, physical world. Surely dualism is the stuff of fantasy, an indulgence of poets and daydreamers, an echo of antiquated worldviews long ago demolished by the relentless progress of science. Though we may occasionally find comfort in imagining that our minds are special, in our more sober moments we must face the facts: our thoughts and feelings, and those of our loved ones, are just as much a part of the brute material order as sticks and stones.

This sentiment expresses a common attitude. The prevalence of this attitude may explain why physicalism, the view that sensations and other mental states are entirely physical, is generally the default position about the mind. On first approaching the mind–body problem, most scientifically minded people assume that physicalism simply has to be true.

However, the sentiment above seriously misrepresents present-day versions of dualism, the belief that some mental states are nonphysical. Many contemporary dualists are fully *naturalistic*. That is, they hold that mental states are just as much a part of the natural order as sticks and stones; and they favor a scientific approach to the mind, one that is independent of religious considerations. In essence, the contemporary dispute between physicalists and naturalistic dualists is a disagreement about what kinds of *data* there are about the nature of mind, and what sort of *theory*—dualist or physicalist—best explains the data.

In this essay, I defend naturalistic dualism. I take, as my starting point, an argument made by René Descartes in his *Meditations*. I expand and defend this argument, drawing on some ideas developed by contemporary philosophers.¹ The expanded argument is, I think, much more powerful than most physicalists recognize. After making my case for dualism, I offer some criticisms of physicalism. The paper will close by defending dualism from the charge that the picture of reality it provides is unacceptably *spooky*.

But first, I must explain in more detail the point at issue between physicalists and dualists. What is it, precisely, that physicalists assert, and dualists deny? Our answer to this question will reveal the type of reasoning that a defense of dualism must employ.

1. PHYSICALISM, DUALISM, AND THE NEED FOR THOUGHT EXPERIMENTS

Physicalism comes in various forms. I will focus on the most straightforward version, the *identity thesis*, which has been defended by numerous philosophers. According to the identity thesis, every type of mental state is identical to some type of physical state. Here is an example of a particular identity claim, relating pain to C-fibers, a group of nerve cells that fire when pain is experienced.

(P) Pain = C-fiber stimulation²

This article was commissioned by the editor expressly for the thirteenth edition of this anthology.
Copyright © Brie Gertler, 2007.

In this statement, “pain” refers to a type of sensation, the type of sensation you have when you stub your toe or bite your lip. This type of sensation is usually caused by tissue damage (in the toe or the lip), but it needn’t be. For instance, a person who has had both feet amputated could nonetheless experience the “stubbed toe” sensation. So damage to the toe is merely the ordinary cause of that sort of sensation; the sensation itself is not spatially located *in* the toe. And it is the sensation itself, not its cause, which is most obviously a *mental* state. Throughout this essay, “pain” will refer to the sensation itself, and not its usual or real cause.

Actually, identity theorists are not committed to any particular identity thesis, such as (P). Their key claim is that every type of mental state is identical to some type of physical state, but they needn’t claim that science has yet uncovered any particular identities. Still, (P) illustrates the *kind* of identity the physicalist has in mind. I will continue to use it in this way.

The first thing to notice about (P) is that it is extremely strong. It does not say merely that pain is perfectly correlated with C-fiber stimulation. Nor does it say that pain will always accompany, and be accompanied by, C-fiber stimulation. Rather, it says that pain just *is*—is nothing over and above—C-fiber stimulation. This means that pain *couldn’t possibly* be present in the absence of C-fiber stimulation, or vice versa, any more than Superman could be present in the absence of Clark Kent, or water could be present in the absence of H₂O. So an identity statement goes beyond a claim about what is *actually* the case, to a claim about what is *possible*—or, really, what is *impossible*.

Because identity statements (statements about what is identical to what) entail that certain scenarios are impossible, they cannot be confirmed by empirical methods alone. To see why this is, imagine that we are in the best-case situation for empirical investigation: We are somehow able to monitor all of the creatures in the universe, and to determine which creatures have C-fibers that are undergoing stimulation, and which creatures are experiencing pain. (To determine whether people are experiencing pain, we might simply ask them; to determine this for

infants and animals, we might observe their behavior.) Suppose that we learn that these are perfectly correlated: C-fiber stimulation is present in all and only those creatures that are currently experiencing pain. Even this very tidy result does not establish the identity thesis, for we still do not know whether it is *possible* that one of these be present in the absence of the other. The perfect correlation might be an odd coincidence. Or, more plausibly, it might be that one of these causes the other. (A prominent dualist position holds that physical events, like the firing of C-fibers, *cause* mental events like pain. This allows us to explain how aspirin can block pain by affecting the brain.) Or it might be that pain and C-fiber stimulation are products of a common cause: Perhaps stubbing one’s toe simultaneously causes both the pain sensation and C-fiber stimulation. These explanations of the correlation are alternatives to saying that pain is *identical to* C-fiber stimulation (e.g., if C-fiber stimulation causes pain then these cannot be identical, for nothing causes itself).

The upshot is that even a perfect correlation does not establish the identity thesis. It shows only that one of these factors *is not*, in fact, present in the absence of the other. But it does not show that one of these factors *could not possibly* be present in the absence of the other. And the identity thesis requires this latter, stronger claim.

If the identity thesis is true, then physicalism is true. Dualists deny the identity thesis, and believe that mental states are distinct from (that is, they are not identical to) physical states. Dualism thus implies that it is *possible* that pain is present in the absence of C-fiber stimulation. Dualists can accept that these are perfectly correlated; but, they claim, these are two distinct states, and so it is not *impossible* that one is present while the other is absent. (Again, perhaps their perfect correlation is due to some causal connection.)

I will defend dualism by arguing that it is possible that you experience pain even if you are in *no* physical state, that is, even if you have no body whatsoever. If pain can occur in the absence of *any* physical state, then physicalism is false.³

Because the dispute between physicalism and dualism concerns whether a particular scenario is

possible, empirical evidence will not resolve it. We must therefore turn to another sort of evidence, gained from thought experiments. Thought experiments can help us to determine whether situations that don't actually occur—such as the presence of pain in the absence of C-fiber stimulation—are nonetheless possible.

Unlike scientific experiments, which are ordinarily conducted in a laboratory, thought experiments are conducted from the comfort of the armchair. One performs a thought experiment by attempting to imagine a given scenario, and then carefully reflecting on the outcome of this exercise. It may seem odd to think that such armchair reflection can yield genuine progress on a thorny issue like the mind–body problem. But because the dispute between physicalists and dualists concerns the sheer *possibility* of a given scenario, neither of these positions can be established without thought experiments.⁴

Here are some examples of thought experiments designed to reveal whether a given scenario is possible.

(A) You have never seen a book that is 100 feet tall. (According to *Guinness World Records*, the world's largest book—a photographic tour of the Asian country of Bhutan—is 5 feet tall by 7 feet wide.) Still, a moment's reflection reveals that you can imagine a 100-foot-tall book. While the high costs of production and the dim sales prospects mean that no one is likely ever to create such a book, these impediments don't absolutely *rule out* its creation. For instance, it is easy to imagine the production being financed by an eccentric millionaire.

We have just resolved the question, "Is it possible that a book be 100 feet tall?" by use of a thought experiment. On reflection, there seemed to be only practical obstacles to the creation of a 100-foot-tall book, and we could imagine overcoming those obstacles. Short of building the giant book ourselves, it is hard to see how this question could be resolved without the use of thought experiments.

(B) You have never seen an object that is blue all over and (simultaneously) orange all over. Can you imagine such an object? It seems clear that you cannot. This exercise of imagination leads

you to believe that nothing could possibly be blue all over and orange all over.

(C) You have never encountered a married bachelor. I expect that you will find yourself unable to conceive of a married bachelor, since you will be unable to conceive of a married person who meets the requirements for being a bachelor (which include, of course, being unmarried). Because nothing can be simultaneously married and unmarried, it is impossible that there be any married bachelors.

In the first two cases, you drew on your *concepts* (**book**, **blue object**, etc.) in trying to *imagine* the described scenario. But in this third case your concepts are more directly involved, because the thought experiment involves *conceiving* rather than *imagining*. To imagine something is to form a sensory (perhaps visual) picture of it. Because no picture of a man will fully express his marital status, mere imagination will not do the trick here. Rather, you exercise the concepts directly, and find that you cannot conceive of a married bachelor because your concept **bachelor** includes the concept **unmarried**, and nothing that satisfies that concept can simultaneously satisfy the concept **married**. (It is no surprise that *conceive* and *concept* have the same Latin root.)

Of course, the results of a thought experiment are not guaranteed to be accurate. Like experiments done in the laboratory, thought experiments can go astray in any number of ways: Your imagination may be limited in a way that makes you unable to imagine a 100-foot-tall book; you may mistakenly think that 'bachelor' refers to any man who lives alone; etc. But as with other types of experiment, these mistakes can be minimized if we take care in designing and performing the experiment. Most importantly, engaging in thought experiments is our only way of determining whether scenarios that haven't actually occurred, and will never actually occur, are nonetheless possible.

In these thought experiments, we try to imagine or conceive a particular scenario, to determine whether it is possible. This exercise is known as a *conceivability test*. There are other types of thought experiments as well, but conceivability tests lie at the heart of most of the

leading arguments for dualism, including the one that I will now present.

2. THE DISEMBODIMENT ARGUMENT

The argument that I will use is a descendant of an argument given by Descartes. I will call it the Disembodiment Argument. Its basic structure is disarmingly simple, but of course the devil—or, as Descartes might say, the malicious demon—is in the details.

The argument centers on the possibility that pain is present in the absence of any physical state.⁵ Arguments for dualism usually focus on pain or other sensations, because the experience of sensations seems to present the greatest challenge to physicalism. Descartes himself believed that thoughts, such as “my senses sometimes mislead me,” posed an equal challenge to physicalism, but contemporary philosophers are divided on that claim. We needn’t concern ourselves with that issue. If our argument succeeds in showing that pain can be present in the absence of any physical state, we will have established dualism, for we will have shown that pain is not identical to anything physical, and thus that at least *some* mental states (viz., pains) are not physical. And that is precisely what the dualist maintains, and the physicalist denies.

As with other thought experiments, this one requires that you actively engage in the exercise of imagining or conceiving. This will unfortunately require a tiny bit of pain. So pinch yourself—lightly!—and, while doing so, put yourself into the position of the “I” in the following line of reasoning.

1. Even though I firmly believe that I have physical features, I can conceive of experiencing *this very pain* while possessing no physical features. In other words, I can conceive of experiencing *this very pain* while disembodied.
2. If I can conceive of a particular scenario occurring, then that scenario is possible.

So,

3. It is possible that *this very pain* occurs in a disembodied being.

4. If *this very pain* was identical to some physical state, then it could not possibly occur in a disembodied being.

So,

5. *This very pain* is not identical to any physical state.

So,

(Conclusion) The identity thesis, which says that every mental state is identical to some physical state, is false.

The major premises in this argument are (1) and (2). Premise (3) follows from these; premise (4) expresses an accepted fact about the nature of identity; and premise (5) follows from (3) and (4).

The best way to target this argument, then, is to deny either (1) or (2). This is precisely what physicalists have done. I will first discuss premise (2), and then take up premise (1).

Premise (2)

The chief criticism of premise (2) is that it involves a kind of intellectual hubris. In conceiving of something, we are simply exercising our *concepts*. And why should we think that our concepts accurately reflect the way the world is? Perhaps we are out of touch with reality, and our concepts don’t correspond to real objects or properties. In that case, the fact that we can conceive of a particular scenario occurring provides no reason to think that that scenario is genuinely possible.

In reply to this criticism, it is crucial to note that *all* of our reasoning—in philosophy and elsewhere—must use some concepts, to define the topic we are investigating. In biology, we begin with some concept of reproduction, which empirical investigation may lead us to refine. In ethics, we begin with some concept of the good, which philosophical reasoning helps us to clarify and develop. And of course physicalists also rely on concepts—including concepts of the physical—in defending their view.

Occasionally, we may find that nothing satisfies a given concept, and so we may abandon investigations relating to it. This is what happened in the case of witchcraft: Most people came to deny that anything in reality corresponded to the concept *witchcraft*, and the study of witchcraft

was replaced by research into superstitions and mental pathologies, phenomena that led to the mistaken belief in witchcraft. But while we must allow for the possibility that we'll refine or even abandon our concepts, concepts are indispensable at the outset of an investigation. For there is no way to proceed with an inquiry unless we have *some* concept of the subject matter we are investigating. The blanket objection that conceivability arguments are illegitimate because they use *our concepts* is, then, misguided.

However, there is a more nuanced version of the worry that premise (2) oversteps our intellectual bounds. Unlike the previous objection, this one acknowledges that we must employ concepts in order to reason at all. And it concedes that thought experiments using simple, straightforward concepts such as **bachelor** can help us to determine what is possible. But it rejects premise (2) on the grounds that it is too general: Not *everything* we can conceive is possible. For some of our concepts are clouded or obscure, and so cannot play the proper role in conceivability tests. In particular, this objection says that our concepts **pain** and **physical** are limited or unclear. Using such faulty concepts, what we can or cannot conceive does not reveal what is or is not possible.

This type of objection was advanced against Descartes' original argument by his contemporary Antoine Arnauld. Arnauld pointed out that a geometry student who hadn't yet encountered the proof of the Pythagorean Theorem could argue as follows:

I can conceive of a right triangle with the following property: the square of its hypotenuse is unequal to the sum of the squares of its other two sides. Therefore, it is possible for a right triangle to have this property.

Clearly, this argument is invalid: The fact that the student could conceive this scenario does not mean that the scenario is possible.

Descartes anticipated Arnauld's objection, and had a ready response. He acknowledged that his argument will fail unless the relevant concepts are "complete and adequate." The student was able to conceive of a right triangle that violated the Pythagorean Theorem only because his concept **right triangle** was limited. But our concepts of

mentality and of the physical are not limited in this way, or so Descartes claimed. That's why he thought that our ability to conceive disembodied pains was genuine evidence that disembodied pains were possible.⁶

In effect, Descartes' reply deflects attention away from premise (2), and makes premise (1) the focus of the argument. For everyone can agree that, if we have a comprehensive understanding of something, then conceivability tests provide a reliable guide to what is possible, involving that thing. It is because you fully understand what it means to be a bachelor that, when you find yourself unable to conceive of a married bachelor, you can justifiably conclude that it's impossible that anything be a married bachelor. And if the geometry student gained a comprehensive grasp of right triangles, then he could similarly rely on conceivability tests. For a comprehensive grasp of right triangles would prevent him from conceiving of a right triangle that violated the Pythagorean Theorem.

Let us agree, then, that conceivability tests can reveal what is possible or impossible, *so long as the concepts involved are sufficiently comprehensive*. This amounts to qualifying premise (2), to read as follows:

- 2.* If, *using concepts that are sufficiently comprehensive*, I can conceive of a particular scenario occurring, then that scenario is possible.

Our new premise (2*) is unobjectionable, for it is trivially true. To say that a concept is "sufficiently comprehensive," in this sense, is just to say that it is suitable for use in conceivability tests. If follows, then, that conceivability tests *using sufficiently comprehensive concepts* reveal what is (and what is not) possible.

Because (2*) is trivially true, the burden of the argument now falls on the first premise. Let us consider premise (1) in more detail.

Premise (1)

We must modify premise (1) to fit our new second premise. It now becomes (1*).

- 1.* *Using concepts that are sufficiently comprehensive*, I can conceive of experiencing *this very pain* while disembodied.

Together, (1*) and (2*) yield (3), so the rest of the argument proceeds as before.

Again because (2*) is trivially true, the entire weight of the argument now rests on (1*). To evaluate (1*), we must determine whether the concepts at work in the “pinch” test are sufficiently comprehensive. That is, when you conceive of experiencing *this very pain* (the pinch) while disembodied, are your concepts **pain** and **physical** sufficiently clear and complete, like your concept **bachelor**? Or might they be confused or incomplete, like the geometry student’s concept **right triangle**?

This is the core issue on which the Disembodiment Argument rests. The argument will succeed only if your concepts of pain and of the physical are sufficiently comprehensive to allow you to conclude, from the fact that you can conceive of experiencing pain while disembodied, that this is possible. Physicalists generally deny this. They believe that, in conceiving of disembodied pain, we are like Arnauld’s geometry student, who conceives of a right triangle that violates the Pythagorean Theorem.

I now turn to discuss our concepts of the physical and of pain. We will see that **pain** plays a much more important role in the Disembodiment Argument.

3. OUR CONCEPT OF THE PHYSICAL

Consider what “physical” means. What conditions must something meet in order to qualify as *physical*?

Descartes thought that the essence of the physical was to be extended in space. But modern physics posits entities—such as fields and waves—that sit uncomfortably with the notion of spatial extension. And because modern physics is still evolving, we should not define the physical in terms of a currently favored theory. For further advances in physics might eventually lead us to reject whatever theory we currently favor.

In response to these worries, many philosophers now conceptualize the physical, for purposes of the mind–body problem, as “the inanimate” or “the nonmental.”⁷ In this way, the concept **physical** is defined by contrast

with the concepts **animate** or **mental**. This suggestion nicely captures the issue at hand. On this reading of “physical,” the central point of contention between physicalists and dualists is as follows:

- Are mental states ultimately, fundamentally nonmental? Are mental states, states of mental things like you and me, ultimately identical (or reducible) to states that are *also* had by inanimate, nonmental things, like sticks and stones?

The dualist will answer “no,” claiming that mental states are part of the basic fabric of the universe. The physicalist will answer “yes,” claiming that mental states are simply patterns of basic physical (that is, nonmental) phenomena. I propose that we adopt this way of conceptualizing the physical. States that are physical, in this sense, are ultimately constituted by phenomena that are nonmental.

As regards the Disembodiment Argument, this understanding of **physical** will affect how you perform the conceivability test described in (1*). To be disembodied is to have *no* physical—ultimately nonmental—features. So you are to attempt to conceive of experiencing *this very pain* while being a *purely* mental entity, an entity that, at least for a moment, has no features in common with sticks and stones.⁸

I now turn to the more significant concept for our purposes, the concept **pain**.

4. OUR CONCEPT OF PAIN

Recall that the Disembodiment Argument will succeed only if our concept of pain is sufficiently comprehensive, in the sense explained above. Is it?

Someone with the concept **pain** might be ignorant of non-essential features of pain: for instance, whether aspirin or ibuprofen more quickly relieves pain. But such features are not relevant to the possibility of disembodied pain. The Disembodiment Argument requires only that we grasp the *essential* features of pain, those that are relevant to the possibility of disembodied pain.⁹

Still, establishing that our concept **pain** is sufficiently comprehensive, in this sense, is a difficult task. The standard expressed by “sufficiently comprehensive” is very high, and as physicalists are

quick to point out, most of our concepts do not meet this standard. The classic example concerns water. Before learning chemistry, each of us could presumably conceive that water was present in the absence of H_2O . And arguably, we can still conceive this. (For we can imagine that chemists develop a new technique for analyzing substances and that, using this technique, they discover that all of the stuff we call "water" contains an additional element, one undetectable by previous methods.) But while we can conceive that water is present in the absence of H_2O , this scenario is of course impossible, because water simply *is* H_2O . Clearly, then, our concept **water** is not sufficiently comprehensive to be used in this type of conceivability test.

Physicalists who cite the water example predict that, once neuroscience has developed more fully, we will come to regard our ability to conceive of disembodied pain in a similar way. That is, we will see that it involves the use of inadequate concepts, and therefore provides no real evidence for the possibility of disembodied pain, just as our ability to conceive of water that is not H_2O provides no evidence for the possibility of non- H_2O water.

To defend the Disembodiment Argument, we need to block the analogy between **water** and **pain**. This requires showing that our concept **pain** is importantly different from concepts such as **water**, to justify the claim that **pain** (unlike **water**) is sufficiently comprehensive. And this is precisely what I will now attempt to do.

First, we must understand *why* our concept **water** is not sufficiently comprehensive. It fails to meet this standard for a simple reason. Namely, we conceptualize water as something that has a *hidden essence*, an essence that can be discovered only by careful scientific investigation. (The hidden essence of water is in fact H_2O .) This is why armchair reflection will not reveal whether non- H_2O water is really possible.

But we don't conceptualize pain this way. *We conceptualize pain as something that has no hidden essence*. Pain wears its essential nature on its sleeve, as it were.

Here we have reached the fundamental, driving idea behind the Disembodiment Argument. As we conceptualize pain, pain has no hidden

essence. If you *feel* that you are in pain, then you are in pain; determining whether you are in pain does not require scientific investigation. For the appearance (feeling) of pain just is pain itself. As one contemporary commentator has put it, "there is no appearance/reality distinction in the case of sensations" such as pain.¹⁰

How does this observation—that the appearance of pain is pain itself—help us to defend (1*)? The challenge, in defending (1*), is to show that our concept **pain** is sufficiently comprehensive for use in conceivability tests. The observation that the *feeling* of pain *is* pain accomplishes this. To determine whether you are really in pain, you need not investigate beyond the feeling of pain. For instance, you need not examine your brain to see whether your C-fibers are undergoing stimulation. (By contrast, investigation *is* required to determine whether something that seems like water—that looks and tastes like water—really is water.) This implies that pain has no hidden essence. So even if pain is perfectly correlated with C-fiber stimulation, C-fiber stimulation is not *essential* to pain; for if it were, then C-fiber stimulation would be the hidden essence of pain, and you couldn't be sure that you were in pain unless you determined that your C-fibers were firing. And if C-fiber stimulation is not essential to pain, then it is not *impossible* that pain be present in the absence of C-fiber stimulation. Hence, pain is not identical to C-fiber stimulation.

Let me be clear about my case here. I am not suggesting that we can't be mistaken about pain.¹¹ My point is just that there is *some* evidence for pain that is absolutely conclusive, for it is, simultaneously, evidence for pain and pain itself. This is evidence of the standard sort: that *hurting* sensation. That sensation is both the appearance of pain and the reality of pain.

Nor am I suggesting that the conceivability test at the heart of the Disembodiment Argument is easy to perform. You must pay close attention to your current "pinching" sensation, making sure that you are not reading more (or less) into it than what it presents. And you must carefully engage in the exercise of conceiving. This is why Descartes advised that, before engaging in the *Meditations*, we should find a quiet place and free our minds from distractions.

But as long as you exercise care, you should be able—through keeping your attention on the pinching feeling, while simultaneously trying to conceive that you are disembodied—to confirm, or disconfirm, (1*) for yourself. For my part, I find that this exercise confirms (1*). And because I assume that we are basically similar, as regards our sensations and concepts, I expect that you will find this as well.

This is, then, the basic argument for dualism. As modified, the key premise of the argument is (1*). That premise is supported by the simple but powerful thought that, according to our concept *pain*, the feeling of pain *is* pain; pain has no essential features that are hidden. So in attending to the pinching sensation, we have access to the essential feature of pain, namely, how it feels. This means that our concept *pain* is sufficiently comprehensive for use in conceivability tests. Your ability to conceive of disembodied pain therefore establishes that disembodied pain is possible, even if (in fact) everyone who experiences pain also has a body. This possibility refutes the identity thesis and establishes dualism.

5. A PHYSICALIST OBJECTION

Physicalists have put forth the following objection to this kind of argument:

The argument shows only that, *according to our concept **pain***, pain may be present in the absence of C-fiber stimulation. Still, we must bear in mind the example of water. *According to our concept **water***, water may be present in the absence of H₂O. Just as the latter case does not lead us to doubt that water = H₂O, the former case should not lead us to doubt that pain = C-fiber stimulation.

As I explained in Section 2, our concepts define the subject matter that we are investigating. And our concepts also determine how we interpret evidence. It is because we conceptualize water as something that has a hidden essence that our discovery that water is correlated with H₂O leads us to believe that water = H₂O. If we conceptualized water differently—e.g., as *any clear liquid*—then we wouldn't accept "water = H₂O."

By the same token, our concept of pain defines what it is that we are investigating when

we try to understand the nature of pain. According to this concept, we are investigating that *hurting* sensation. And because pain has no hidden essence, according to our concept, the discovery that pain is correlated with C-fiber stimulation will not justify "pain = C-fiber stimulation."

Above, I said that empirical investigation could not establish that a certain scenario is *impossible*. This is why even a perfect correlation between pain and C-fiber stimulation could not establish that it is impossible that pain be present without C-fiber stimulation, and hence could not establish that these were identical. How, then, did empirical investigation establish that water = H₂O? The answer is simple: Empirical investigation did *not* single-handedly establish this. Empirical investigation established only a correlation between water and H₂O. What justified the identity claim was an additional fact, a fact about our *concept* of water, namely, that we conceptualize water as having a hidden essence. But we don't take the fact that a substance has a particular appearance (it is clear, drinkable, etc.) to be *conclusive evidence* that it is water; Rather, we think that, to conclusively determine whether something is water, we must discover its microstructure. The fact that water has a hidden essence was already implicit in our concept *water*; and this is why the discovery that water was perfectly correlated with H₂O led us to believe that H₂O was the hidden essence of water, and hence that water = H₂O.

By contrast, we don't think of pain as having a hidden essence. This is why you treat the *appearance* of pain—that *hurting* sensation—as conclusive evidence that you are in pain. And it is why we don't need scientific investigation to discover the *essence* of pain, though of course such investigation can discover interesting and important facts about pain.

Above, I said that concepts are crucially important because they define the topic that we are investigating. Anyone who interprets the correlation between pain and C-fiber stimulation as evidence that pain = C-fiber stimulation is straying from our concept, and hence simply changing the topic. The topic we're investigating, pain, is defined by our concept *pain*; and according to that concept, pain has no hidden essence.

But why not simply change the topic, and claim that nothing corresponds to our original concept *pain*, because every mental state has a hidden physical essence? The problem with this move is that we can still ask why we should accept the claim that *pain* = C-fiber stimulation. In the case of other identities, such as *water* = H_2O , there is a simple answer: You conceptualize water as having a hidden essence, and empirical investigation has discovered that H_2O is that hidden essence. Because there is no parallel justification for "*pain* = C-fiber stimulation," we have no reason to accept that identity claim.

Arguably, this distinctive feature of our pain concept—that, as we conceptualize pain, the appearance of pain *is* its reality—is possessed only by mental concepts. For our concepts of nonmental things do attribute a gap between whether the thing appears to satisfy the concept and whether it actually does so. If this is correct, our argument for dualism cannot be applied to anything that is not mental.

6. BUT WHAT ABOUT MENTAL CAUSATION?

The previous section discussed a physicalist objection to the method used in the Disembodiment Argument. However, a physicalist might instead criticize the *conclusion* of that argument, as follows:

The Disembodiment Argument does provide *some* reason to think that disembodied pain is possible. But we should not embrace this conclusion until we carefully examine its consequences. And in fact the conclusion has one consequence that is utterly unacceptable: if dualism is true, then our mental states do not truly cause our bodily movements. That is, dualism leads to *epiphenomenalism*, the claim that the mental has no physical effects. But epiphenomenalism is repugnant.¹² So even if we are not sure quite where the argument goes wrong, it must go wrong somewhere.

Leading arguments for physicalism rest on the idea that accepting the identity thesis is the only way to preserve mental causation. But only a very strong version of the identity thesis will salvage mental causation as intended; and that version of the thesis faces problems that are as

serious as the worry about epiphenomenalism. Or so I will argue.

The identity thesis is usually motivated by the following idea. If we want to explain why I leaped out of my chair (a physical event), we need not invoke anything other than previous physical events: I sat on a thumb tack, and this caused (via a train of other physical events) C-fiber stimulation, which then caused my leap out of the chair. To say that such an explanation requires us to invoke a mental state like pain *in addition* to C-fiber stimulation would be to deny that physics is complete. But all of our evidence is that physics *is* complete: We have never encountered a case where we were *forced* to go beyond the realm of the physical to explain a physical event. Now if physics is complete, then citing physical events will suffice to explain my leaping out of the chair. We need not mention pain in this explanation. So *if* pain causally contributes to my leaping from the chair, then pain must be physical—that is, it must be *identical* to a physical state such as C-fiber stimulation.

There are various objections that could be made to this argument. To my mind, the most promising is this: The nature of causation, and what it means to say that one event caused another, is far from clear. For example, David Hume argued that the best we can expect, in attempting to explain why an event occurred, is to identify a regularity: *This type of event is regularly preceded by that type of event*. But then why not say that *suddenly leaping out of one's chair* is regularly preceded by *feeling pain*? The physicalist will claim that there is also a physical event that regularly precedes such leaping, and so the mental event is superfluous. However, this response assumes that, in a competition between mental and physical causes, the physical cause will always win out. (Why not say that it is the physical event that is superfluous?) This assumption would be justified if we had some *deeper* account of how physical causation occurs, one that favored physical-to-physical causation over mental-to-physical causation. But there is no accepted understanding of how causation occurs; and contemporary Humeans believe that there is no deeper *fact* about how causation occurs, because causal statements merely report regularities. The bottom line

is this: Physical causation is largely a mystery, and so we should be hesitant to use our limited understanding of physical causation to rule out the possibility of nonphysical causation. The argument that dualism commits us to epiphenomenalism is far from conclusive.

But even if you think that the identity thesis better accommodates mental causation, it faces another problem, sometimes called the *problem of chauvinism*. If pain = C-fiber stimulation, then creatures whose physical structure differs significantly from ours cannot feel pain. In other words, the identity thesis is chauvinistic, because it identifies pain with a physical state that may be specific to creatures on Earth.

To see the problem, suppose that we encounter an alien civilization that seems very similar to ours. In this civilization, legislative debate is often reduced to partisan squabbles. Its members spend their leisure time in a range of activities, from hiking to watching reality TV. It seems easy to imagine that aliens whose activities were so similar to ours could nonetheless differ physiologically from humans. Perhaps their planet does not have the abundance of H₂O and oxygen that Earth has, and so their physical and neural constitution is very different from ours. Even so, it seems entirely conceivable that such creatures feel pain. (We can suppose that some of these aliens are in hospitals, recovering from surgery.) But if pain is *just* C-fiber stimulation, then no creature who lacked C-fibers *could possibly* feel pain. The identity thesis does seem to be objectionably chauvinistic.

To my mind, it is highly implausible that the capacity for pain requires a physical constitution like ours. Remember, what is at issue here is not the *likelihood* that there are alien creatures who feel pain; it is the sheer *possibility* of this scenario. And it seems incredible that pain *couldn't possibly* be experienced by creatures who lacked C-fibers.

I have argued that, while the concern about whether dualism can do justice to mental causation is legitimate, the reasoning to show that dualism commits us to epiphenomenalism is open to question. Accepting the identity thesis obviously commits us to chauvinism, but it is not entirely clear that accepting dualism commits us to epiphenomenalism. And chauvinism is at least as repugnant as epiphenomenalism.

Some physicalists have responded to the chauvinism problem by retreating from the simple version of the identity thesis. In its simple version, the identity thesis claims that each type of mental state is identical to some type of physical state. The retreat from this simple version claims that pain *in humans* is identical to C-fiber stimulation; pain *in aliens* may be identical to some other physical state.¹³ This move does avoid the chauvinism problem. But the retreat from the simple identity thesis means that we cannot explain such seemingly obvious causal generalizations about pain as “being in pain causes one to urgently try to change one’s situation.” For there is no single property that answers to “being in pain”; there is *pain-in-humans*, and *pain-in-aliens*, but on the proposal being considered, these are completely different properties. Because this modification of the identity thesis prevents us from allowing for familiar causal generalizations, it limits the alleged causal benefits of physicalism.

Until this point, I have been concerned solely with “identity” versions of physicalism. But in the face of the difficulties I have described, some physicalists adopt a weaker position, claiming that the relations between physical and mental states fall short of identity. Unfortunately, this weaker, “nonreductive” brand of physicalism faces precisely the same worries about epiphenomenalism that dualism allegedly faces. For the epiphenomenalism worry stems from the claim that pain $\neq$ C-fiber stimulation. So long as my pain isn’t identical to the firing of my C-fibers, then we can always ask *which* of these factors caused me to leap out of the chair. Was it the pain, or was it the C-fiber stimulation? The danger is that the physical factor will fully explain my leap, and thus the mental factor will be superfluous.¹⁴ The benefit of the identity thesis is that it blocks such questions. (Compare: Because Superman is identical to Clark Kent, it makes no sense to ask whether it was Superman *or* Clark Kent who stepped out of the phone booth.) Because avoiding epiphenomenalism is the chief consideration in favor of physicalism, these weaker physicalist positions seem unmotivated. Finally, it’s worth noting that these weaker brands of physicalism are just as vulnerable to the Disembodiment Argument,

for even the weakest brand of physicalism must hold that you could not feel the pinch you're feeling right now if you were disembodied.

Let me sum up the results of this section. According to physicalists, dualism implies that mental events, such as thoughts and sensations, never have physical effects. Many physicalists claim that the only way to avoid this "repugnant" epiphenomenalism is to accept the identity thesis. In response, I have outlined an objectionable consequence of the identity thesis: that those with different physical constitutions could not possibly experience pain or any other sensation that we experience. This chauvinistic result is at least as worrisome as epiphenomenalism. And the simple identity thesis is patently chauvinistic, while it is less clear that dualism leads to epiphenomenalism. Chauvinism can be avoided by modifying the identity thesis, but this modification sacrifices much of the alleged causal benefits of the identity thesis.

7. BUT ISN'T DUALISM SPOOKY?

Despite all this, there may be a lingering sense that dualism is just plain weird, or *spooky*.

But is dualism any spookier than physicalism? Bertrand Russell observed that those who accept the findings of modern science are hardly in a position to accuse dualists of spookiness.

The plain man thinks that material objects must certainly exist, since they are evident to the senses. Whatever else may be doubted, it is certain that anything you can bump into must be real; this is the plain man's metaphysic. This is all very well, but the physicist comes along and shows that you never bump into anything: even when you run your hand along a stone wall, you do not really touch it. When you think you touch a thing, there are certain electrons and protons, forming part of your body, which are attracted and repelled by certain electrons and protons in the thing you think you are touching, but there is no actual contact. . . . The electrons and protons themselves, however, are only crude first approximations, a way of collecting into a bundle either trains of waves or the statistical probabilities of various different kinds of events. Thus matter has become altogether too ghostly to be used as an adequate stick with which to beat the mind.¹⁵

Russell's point is that commitments of physical science, which is unquestionably naturalistic, are just as spooky or "ghostly" as dualism. And Russell was writing in 1935; subsequent advances in physics have made the physical realm appear all the more exotic. Contemporary physicists do not agree on the basic structure of the world, but the entities posited by relatively mainstream physical theories—vibrating strings, basic particles subject to quantum indeterminacy—seem spooky enough.

Surely, dualism does not *spookify* our overall picture of the world. That picture was already plenty spooky, due to the progress of physics itself. So there is no reason to hold that dualism is spooky, or that it conflicts with a broadly scientific, naturalistic picture of reality.

Conclusion

Most contemporary thinkers believe that minds are part of the natural order, and that scientific research can yield important information about how the mind works. Naturalistic dualists agree. And dualism is not undermined by empirical evidence, because empirical evidence reveals only correlations. Moreover, dualists can accept causal explanations of these correlations (e.g., the correlation between taking aspirin and pain reduction is explained by the fact that aspirin suppresses the biochemicals responsible for activating C-fibers).

At the outset of this essay, I noted that the dispute between physicalists and dualists can be interpreted as a disagreement as to what kinds of data there are about the nature of mind, and what kind of theory will best explain those data. The thought experiment at the center of the Disembodiment Argument provides data that favor dualism: We seem able to conceive of disembodied pain. Given that our concept of pain appears to be sufficiently comprehensive for use in a conceivability test, the proper conclusion is that pain could be present in a being that lacked physical features. This means that dualism is true.

Physicalists will discount the importance of data generated by conceivability tests. But these data are essential to defining the topic at hand, whether that topic is bachelors, water, or pain. So any claim that clashes with these data diverges

from our original topic, and should therefore be rejected. The claim that “pain = C-fiber stimulation” clashes with the data generated by our conceivability test. It thus diverges from our original topic—pain, as defined by our concept *pain*—and hence we should reject it.

Any naturalist worth her salt will follow the evidence where it leads. And in this case, the evidence leads to dualism.

NOTES

1. My development of this argument most closely parallels Saul Kripke's argument in Lecture III of his book *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980). The other leading contemporary arguments for dualism are also naturalistic: Frank Jackson's Knowledge Argument, given in “The Qualia Problem,” in this volume; and David Chalmers' Zombie Argument, given in Chapter 4 of his book *The Conscious Mind* (Oxford: Oxford University Press, 1996). Some of the ideas in this essay borrow from these other arguments as well.
2. This is a *type-identity* thesis: it identifies a *type* of mental state with a *type* of physical state. For simplicity, I will focus on type-identities. But my argument also challenges the thesis that every *particular* (or “token”) mental state, such as the pain I'm feeling right now, is identical to some *particular* (“token”) physical state. This latter claim is known as a *token-identity* thesis. Unlike the type-identity thesis, it allows that two instances of a single mental type (such as pain) can belong to different physical types. This distinction will surface briefly in Section 6.
3. Some philosophers believe that conceivability arguments in this direction don't succeed in undermining physicalism, since my ability to conceive of disembodied pain shows only that such a pain *could* be disembodied, not that my pain is actually distinct from my body. But we can sidestep this objection by focusing, in the thought experiment, on the idea that *this very pain* could be disembodied.
4. Of course, empirical evidence can tell us that a certain scenario is *possible*. If we observed that pain occurred in the absence of C-fiber stimulation, we would of course conclude that this was possible. So it might seem that this type of empirical observation could support dualism. However, even this observation would not settle the issue. For the physicalist will simply take this as evidence that we must look elsewhere to find the physical state identical to pain. Unless we have empirical evidence of a pain that is unaccompanied by *any* physical state—which seems highly unlikely—the physicalist can always hold out for the discovery of a physical state that does perfectly correlate with pain. I will assume, for the sake of argument, that there is a physical state that perfectly correlates with pain, and I will follow standard practice in speaking as if C-fiber stimulation is this state.
5. This means that it applies both to identities like “pain = C-fiber stimulation” and to the claim that every particular pain is identical to *some* physical state or other. That is, it applies to both type-identity and token-identity versions of physicalism. (These versions were distinguished in note 3.)
6. In his version of the argument, Descartes doesn't distinguish pains from thoughts more generally.
7. David Papineau proposes that we understand the physical as “the inanimate.” See Papineau, *Thinking about Consciousness*. Oxford University Press, 2002. Barbara Montero proposes that we understand it as “the nonmental.” See “The Body Problem,” *Noûs* 33 (1999), pp. 183–200.
8. This should be qualified a bit, for purely mental entities may share temporal features with sticks and stones, e.g., existing on a Tuesday. And of course they may share necessary features like “being such that $2 + 2 = 4$.” So the issue is whether you can conceive of experiencing *this very pain* while sharing no contingent, nontemporal features with inanimate, nonmental objects.
9. Strictly speaking, our concept *pain* is sufficiently comprehensive only if we grasp both the essential features of pain and any features entailed by those essential features. The geometry student erred because he didn't grasp a property (the property expressed by the Pythagorean Theorem) entailed by the essential feature of a right triangle *being a three-sided polygon with one angle equal to 90°*.
10. Christopher Hill, *Sensations: A Defense of Type Materialism* (Cambridge: Cambridge University Press, 1991), p. 127.
11. You might mistakenly believe that you're in pain if a trusted friend tells you that you are, at a time when you are steadfastly refusing to introspect your own feelings. (This is certainly an odd case, however, and seems more plausible for emotional pain than for the kind of pinch we are concerned with.) Or you might be tricked into believing that you're in pain: In a (cruel) fraternity hazing ritual, blindfolded recruits are told that they will be burnt with a cigarette in a certain spot. Ice is

- then applied to the spot, yielding shock rather than pain. Still, for a moment the victims react as if they were in pain, and they seem to believe that they are undergoing pain. (An alternative interpretation is that the recruits *are* in pain, but the pain has a curious origin: It is produced by the expectation of pain.)
12. The charge that epiphenomenalism is "repugnant" was made by David Lewis in "What Experience Teaches." This article appeared in W. Lycan, ed., *Mind and Cognition* (Oxford: Blackwell Publishing, 1990), pp. 499–519.
 13. Lewis, "Mad Pain and Martian Pain," in N. Block, ed., *Readings in the Philosophy of Psychology*, Vol. I. (Harvard University Press, 1980), pp. 216–222.
 14. The idea that non-reductive physicalism faces the problem of epiphenomenalism was made by Jaegwon Kim. See "The Nonreductivist's Troubles with Mental Causation," reprinted as Chapter 17 of Kim's book *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993).
 15. Russell, "What Is the Soul?" from his book *In Praise of Idleness and Other Essays* (NY: W. W. Norton, 1935).

The Qualia Problem

FRANK JACKSON

Frank Jackson has written many works in metaphysics, ethics, and the philosophy of mind. He teaches at the Princeton University.

I am what is sometimes known as a "qualia freak." I think that there are certain features of the bodily sensations especially, but also of certain perceptual experiences, which no amount of purely physical information includes. Tell me everything physical there is to tell about what is going on in a living brain, the kind of states, their functional role, their relation to what goes on at other times and in other brains, and so on and so forth, and be I as clever as can be in fitting it all together, you won't have told me about the hurtfulness of pains, the itchiness of itches, pangs of jealousy, or about the characteristic experience of tasting a lemon, smelling a rose, hearing a loud noise, or seeing the sky.

There are many qualia freaks, and some of them say that their rejection of Physicalism* is an unargued intuition. I think that they are being unfair to themselves. They have the following argument. Nothing you could tell of a

physical sort captures the smell of a rose, for instance. Therefore, Physicalism is false. By our lights this is a perfectly good argument. It is obviously not to the point to question its validity, and the premise is intuitively obviously true both to them and to me.

I must, however, admit that it is weak from a polemical point of view. There are, unfortunately for us, many who do not find the premise intuitively obvious. The task then is to present an argument whose premises are obvious to all, or at least to as many as possible. This I try to do with what I will call "the Knowledge argument."

In the final section I tackle the question of the causal role of qualia. The major factor in stopping people from admitting qualia is the belief that they would have to be given a causal role with respect to the physical world and especially the brain; and it is hard to do this without

Reprinted by permission of John Wiley & Sons, Ltd.

*The theory that all information is ultimately physical information—information about physical states and physical events, couched in exclusively physical language—Eds.