

follows: These identities *enable* desirable psychological explanations while *disabling* the improper demands for explanation of psychoneural correlations.<sup>13</sup>

How good is this argument? Unfortunately, not very good: The argument turns out to be problematic, for reasons similar to those that made explanatory argument I questionable. The trouble is that in both arguments the identities in question do not seem to do any explanatory work and hence are not qualified to benefit from the principle of inference to the best explanation. We can accept the claim that derivation ( $\theta$ ) gives a neurophysiological explanation of why pain causes distress: Laws of neurophysiology directly explain why Cfs causes neural state N, and given the identities “pain = Cfs” and “distress = neural state N,” we would be justified in claiming that neurophysiological laws explain the fact that pain causes distress. This is so because, given the two identities, the statements “pain causes distress” and “Cfs causes neural state N” state one and the same fact. There is here one fact described in two ways—in the vernacular vocabulary and in the scientific vocabulary.

This shows just what goes wrong with explanatory argument II: The identities “pain = Cfs” and “distress = neural state N” do *no explanatory* work in this derivation. Their role is to enable us to *re-describe* a fact that has already been explained. The explanatory activity is over and finished at the second line when “Cfs causes neural state N” has been derived from, and thereby explained by, laws of neurophysiology. What the identities do is allow us to *rewrite* “Cfs causes neural state N” as “pain causes distress,” by putting equals for equals. This is useful in presenting our explanatory accomplishment in neuroscience in the familiar “folk” language, but this involves no *explanatory* activity. The verdict, therefore, seems inescapable: Since the psychoneural identities have no involvement in explanation, they are ineligible as beneficiaries of the principle of inference to the best explanation. If there is a beneficiary of this principle in this situation, it is the laws of neuroscience because they do the explanatory work!

Our conclusion, therefore, has to be that both forms of the explanatory argument are vulnerable to serious objections. Their shared weakness is a lack of clear appreciation of just what role the psychoneural identities play in the explanations in which they supposedly figure. Our main contention has been that both arguments invoke, but misapply, the rule of inference to the best explanation, a principle that itself is far from uncontroversial.

## AN ARGUMENT FROM MENTAL CAUSATION

By mental causation we mean any causal relation involving a mental event. A pin is run into your palm, causing you a sharp pain. The sudden pain causes

you to cry out and quickly pull back your hand. It also causes a feeling of distress and a desire to be rid of it. Causal relations involving mental and physical events are familiar facts of our everyday experience.

But pains do not occur without a physical basis; let us assume that pains are lawfully correlated with neural state N. So the sharp pain that caused the withdrawal of your hand has an occurrence of N as its neural substrate. Is there any reason for not regarding the latter, a neural event, as a cause of your hand's jerky motion?

Suppose we try to trace the causal chain backward from your hand's movement. The jerky motion was presumably caused by the contraction of muscles in your arm, which in turn was caused by neural signals reaching the muscles. The movement of neural signals is a complex physical process involving electrochemical interactions, and if we keep tracing the series of events backward to its source, we can expect it to culminate in a region in the central nervous system, perhaps in the cortex. Now ask yourself: Will this chain ever reach, or go through, a mental experience of pain, the pain you experienced when the pin was stuck in your palm? What could the transition from a neural event to a nonphysical, private pain event be like? Or the transition from a private pain experience to a public physicochemical neural event? How can a pain experience affect the motion of even a single molecule—speeding it up or slowing it down, or changing its direction? How can that happen? Is it even conceivable? It boggles our imagination!

The chances are that the causal chain culminating in your hand's jerky movement, when traced backward, will completely bypass your pain; there will be more and more neural-physical events as you keep going back, but no mental experiences. Nor does it make sense to postulate a purely mental causal chain, independent of the neural-physical chain, somehow reaching your muscles. (That's known as telekinesis—an alleged "psychic" phenomenon involving a mind causing a physical change at a distance, like bending a spoon by intensely gazing at it.) It seems, then, that the only way to salvage the pain as a cause of your hand motion is to think of it as a neural event. Which neural event? The best and most natural choice is its neural substrate, N (as we supposed), the state that is necessary and sufficient for the occurrence of the pain. This in brief is the causal argument, somewhat informally presented, for identifying mental states, especially states of consciousness, with neural states.

There is a more systematic, and currently influential, version of the causal argument that will now be presented. It begins with a premise asserting that mental causation is real:

- i. Mental phenomena have effects in the physical world.

In this context, we take (i) as uncontroversial. Our beliefs and desires surely have the power to move our limbs and thereby enable us to cause things around us to be rearranged—moving the books from my desk to the bookshelves, emptying a waste basket, digging my car out of a snowbank, and starting an avalanche. If our mental states had no causal powers to affect physical things and events around us, we would cease to be agents, only helpless spectators of the passing scene. If that were true, our self-conception of ourselves as effective agents in the world would suffer a complete collapse.

Here is the second premise:

- ii. [The causal closure of the physical domain] The physical world is causally closed. That is, if any physical event is caused, it has a sufficient physical cause (and a wholly physical causal explanation).

According to this principle, the physical world is causally self-contained and self-sufficient. It doesn't say that every physical cause has a sufficient physical cause—that is the principle of physical causal determinism. So (ii) is compatible with indeterminism about physical events. What (ii) says is that for any physical event, if we were to trace its causal ancestry, this need never take us outside the physical world. If a physical event has no physical cause, then it has no cause at all and no causal explanation. Further, this principle is compatible with dualism and other forms of nonphysicalism: As far as it goes, there could be a Cartesian world of immaterial minds, alongside the physical world, and all sorts of causal relations could hold in that world. The only thing, according to physical causal closure, is that the physical world must be causally insulated from such worlds; there can be no injection of causal influence into the physical world from outside. This means that there can be no “miracles” brought about by some transcendental, supernatural causal agents from outside physical space-time.

On Descartes's interactionist dualism, the physical causal closure fails: When an immaterial soul makes the pineal gland vibrate, thereby setting in motion a chain of bodily events, the motion of the pineal gland is caused, but it has no physical cause and no physical explanation. And this means that our physical theory would remain forever incomplete in the sense that there are physical events whose occurrences cannot be physically explained. A complete theory of the physical world would require references to nonphysical, immaterial causal agents and forces.

Why should we accept the causal closure of the physical domain? We will enumerate some reasons here without going into great detail.<sup>14</sup> First, there is

the widely noted success of modern science, in particular theoretical physics, which we take to be our basic science. Physics is all-encompassing: Nothing in the space-time world falls outside its domain. If a physicist encounters a physical event for which there is no ready physical explanation, or physical cause, she would consider that as indicating a need for further research; perhaps there are as-yet undiscovered physical forces. At no point would she consider the possibility that some nonphysical force outside the space-time world was the cause of this unexplained physical occurrence. The same seems to be true of research in other areas of science—broadly physical science including chemistry, biology, geology, and the like. If a brain scientist finds a neural event that is not explainable by currently known facts in neural science, what is the chance that she would say to herself, “Maybe this is a case of a Cartesian immaterial mind interfering with neural processes, messing up my experiment. I should look into that possibility!” We can be sure that would never happen. What would such research, investigating the workings of immaterial souls, look like? Where would you start? It isn’t just that the principle of physical causal closure is the operative assumption in scientific research—remember that in science success is what counts. It may well be that there is a conceptual incoherence in the idea that there are nonphysical causal forces outside space-time that can causally intervene in what goes on in the space-time world.<sup>15</sup>

From these two premises, (i) and (ii), we have the desired conclusion:

(i) Mental phenomena are physical phenomena.

You might point out, rightly, that the only proposition we are entitled to derive is that only those mental phenomena that cause physical events are physical events.<sup>16</sup> Strictly speaking, that is correct, but remember this: Causation is transitive—that is, if one event causes another, and this second event causes a third, then the first event causes the third. If a mental event causes another mental event, which causes a physical event, the first event causes this physical event, and our argument pronounces it to be a physical event. Such chains of mental events can be as long as you wish; as long as a single event in this chain causes a physical event, every event preceding it in the chain qualifies as a physical event. This should pretty much cover all mental events; it is hard to imagine a mental causal chain consisting exclusively of mental events not touching anything physical anywhere. Even if there were such exceptions, the main physicalist point is made. A qualified conclusion stands: Mental events that have effects in the physical domain are physical events. The pain that causes your hand to pull back in a jerky motion and makes you cry

“Ouch!” is a physical event. But which physical event? What better candidate is there than the brain state that is the neural correlate of pain, namely Cfs? Cfs is a necessary and sufficient condition for the occurrence of pain, and it occurs exactly at the same time as the pain.

If in spite of these considerations you still want to insist on the pain as a separate cause of the hand movement, think of a new predicament in which you will find yourself. For the hand movement would now appear to have two distinct causes, the pain and its neural correlate Cfs, each presumably sufficient to bring it about. Doesn't that make this (and every other case of mental-to-physical causation) a case of causal overdetermination, an instance in which two independent causes bring about a single effect? Given that the hand withdrawal has a sufficient physical cause, namely Cfs, what *further* causal contribution can the pain make? There seems no leftover causal work that the pain has to be called on to perform. Again, the identification of the pain with Cfs appears to dissolve all these puzzles. There is, of course, the epiphenomenalist solution: Both the hand withdrawal and the pain are caused by Cfs, and the pain itself has no further causal role in this situation. But unlike the identity solution, the epiphenomenalist move renders the pain causally inert and ends up rejecting our initial assumption that a sharp pain caused the hand's jerky motion.

Perhaps a reconsideration of that assumption may be in order. The identification of a conscious pain experience with some molecular physical processes in the brain strikes some people as totally incredible and still others as verging on incoherence. If given a choice between taking pain and other experiences as physical processes in the brain on one hand and their causal impotence on the other, some may well consider the latter a preferable option. At this point, what the causal argument does is to give us a choice between psychoneural identity and epiphenomenalism: If you want to protect mental events from epiphenomenalism, you had better identify them with physical processes in the brain. To some, this may seem tantamount to discarding what is distinctively mental in favor of molecular physical processes in the body. On the other hand, if you are unwilling to embrace psychophysical identity, you put the causal powers of mentality in jeopardy. What good is our mentality if it is epiphenomenal? We will return to some of these issues later (chapter 7).

## AGAINST PSYCHONEURAL IDENTITY THEORY

There are three main arguments against the mind-brain identity theory. They are the epistemological argument, the modal argument, and the multiple realization argument. We consider each in turn.