Some ancient Greeks thought that the heart was the organ responsible for thoughts and feelings—an idea that has survived, we are told, in the traditional symbolism of the heart as signifying love and romance. But the Greeks got it wrong; we now know, as surely as such things can be known, that the brain is where the action is as far as our mental life is concerned. If you ask people where their minds or thoughts are located, they will point to their heads. Does this mean only that the mind and brain share the same location, or something stronger, namely, that the mind is the brain? We consider here a theory that advocates this stronger claim—that the mind is identical with the brain and that for a creature to have mentality is for it to have a brain with appropriate structure and capacities.

**Mind-Brain Correlations**

But what makes us think that the brain is “the seat of our mental life,” as Descartes might have put it? The answer seems clear: There are *pervasive and systematic psychoneural correlations*, that is, *correlations between mental phenomena and neural states of the brain*. This is not something we know a priori; we know it from empirical evidence. We observe that injuries to the brain often have a dramatic impact on mental life, affecting the ability to reason, recall, and perceive, and that they can drastically impair a person’s cognitive capacities and even alter her personality traits. Chemical changes in the brain brought on by ingestion of alcohol, antidepressants, and other psychoactive drugs affect our moods, emotions, and cognitive functions. When a brain concussion
knocks us out, our conscious life goes blank. Sophisticated brain imaging techniques allow us to "see" just what is going on in our brains when we are engaged in certain mental activities, like seeing green or feeling agitated. It is safe to say that we now have overwhelming scientific evidence attesting to the centrality of the brain and its activities as determinants of our mental life.

A badly scraped elbow can cause you a searing pain, and a mild food poisoning is often accompanied by stomachaches and queasy feelings. Irradiations of your retinas cause visual sensations, which in turn cause beliefs about objects and events around you. Stimulations of your sensory surfaces lead to sensory and perceptual experiences of various kinds. However, peripheral neural events are only remote causes; we think that they bring about conscious experiences only because they cause appropriate states of the brain. This is how anesthesia works: If the nerve signals coming from sensory peripheries are blocked or the normal functions of the brain are interfered with so that the central neural processes that underlie conscious experience are prevented from occurring, there will be no experience of pain—perhaps no experience of anything. It is plausible that everything that occurs in mental life has a state of the brain (or the central nervous system) as its proximate physical basis. It would be difficult to deny that the very existence of our mentality depends on the existence of appropriately functioning neural systems: If all the cells and molecules that make up your brain were scattered in intergalactic space, your whole mental life would vanish at that moment, just as surely as annihilating all the molecules making up your body would mean its end. At least that is the way things seem. We may summarize this in the following thesis:

Mind-Brain Correlation Thesis. For each type M of mental event that occurs to an organism o, there exists a brain state of kind B (M’s “neural correlate” or “substrate”) such that M occurs to o at time t if and only if B occurs to o at t.

According to this thesis, then, each type of mental event that can occur to an organism has a neural correlate that is both necessary and sufficient for its occurrence. So for each organism there is a set of mind-brain correlations covering every kind of mental state it is capable of having.

Two points may be noted about these brain-mind correlations:

1. They are “lawlike”: The fact that pain is experienced when certain of your neurons (say, C-fibers and Aδ-fibers) are activated is a matter of lawful regularity, not accidental, or coincidental, co-occurrences.
2. Even the smallest change in your mental life cannot occur unless there are some specific (perhaps still unknown) changes in your brain state; for example, when your headache goes away, there must be an appropriate change in your neural states.

Another way of putting these points, though this is not strictly equivalent, is to say that mentality supervenes on brain states. Remember that this supervenience, if it indeed holds, is something we know from observation and experience, not a priori. Moreover, specific correlations—that is, correlations between specific types of mental states (say, pain) and specific types of brain states (say, the activation of certain neural fibers)—are again matters of scientific research and discovery, and we may assume that many of the details about these correlations are still largely unknown. However, it is knowledge of these specific correlations, rough and incomplete though it may be, that ultimately underlies our confidence in the general thesis of mind-brain correlation and mind-brain supervenience. If Aristotle had been correct (and he might have been correct) about the heart being the engine of our mentality, we would have a mind-heart correlation thesis and mind-heart supervenience, instead of the mind-brain correlation thesis and mind-brain supervenience.

**Making Sense of Mind-Brain Correlations**

When a systematic correlation between two properties or types of events has been observed, we want an explanation, or interpretation, of the correlation: Why do the properties F and G correlate? Why is it that an event of type F occurs just when an event of type G occurs? We do not want to countenance too many “brute,” unexplained coincidences in nature. An explanatory demand of this kind becomes even more pressing when we observe systematic patterns of correlation between two large families of properties, like mental and neural properties. Let us first look at some examples of property correlations outside the mind-brain case:

a. Whenever the ambient temperature falls below 20 degrees Fahrenheit and stays there for several days, the local lakes and ponds freeze over. Why? The answer, of course, is that the low temperature causes the water in the ponds to freeze. The two events are causally related, and that is why the observed correlation occurs.

b. You enter a clock shop and find an astounding scene: Dozens and dozens of clocks of all shapes and sizes are busily ticking away, and
they all show exactly the same time, 2:00. Awhile later, you see all of
them showing exactly 2:30, and so on. What explains this marvelous
correlation among these clocks? It could not be a coincidence, we
think. One possible answer is that the shopkeeper synchronized all
the clocks, which are all working properly, before the shop opened
in the morning. Here, a *common cause*, the shopkeeper’s action in
the morning, explains the correlations that are now observed; to put
it another way, one clock showing 3:30 and another showing the
same time are *collateral effects of a common cause*. There are no di-
rect causal relationships between the clocks that are responsible for
the correlations.

**c.** We can imagine a slightly different explanation of why the clocks
are keeping the same time: These clocks actually are not very accu-
rate, and some of them gain or lose time markedly every five min-
utes or so. But there is a little leprechaun whose job is to run around
the shop, unseen by the customers, synchronizing the clocks every
minute. That is why every time you look, the clocks show the same
time. This again is a *common-cause* explanation of a correlation, but
it is different from the story in (b) in the following respect: This
explanation involves a continued intervention of a causal agent,
whereas in (b) a single cause in the past is sufficient. In neither case,
however, is there a direct cause-effect relationship between the correlated events.

**d.** Why do temperature and pressure covary for gases confined in a
rigid container? The temperature and pressure of a gas are both de-
pendent on the motions of the molecules that compose the gas: The
temperature is the average kinetic energy of the molecules, and the
pressure is the momentum imparted to the walls of the container
(per unit area) by the molecules colliding with them. Thus, the rise
in temperature and the rise in pressure can be viewed as *two aspects*
of one and the same underlying microprocess.

**e.** Why does lightning occur just when there is an electric discharge be-
tween clouds or between clouds and the ground? Because lightning
simply is an electric discharge involving clouds and the ground.
There is here only one phenomenon, not two that are correlated with
each other, and what we thought were distinct correlated phenom-
ena turn out to be one and the same event, under two different de-
scriptions. Here an apparent correlation turns out to be an *identity*. 
f. Why do the phases of the moon (full, half, quarter, and so on) co-vary with the tidal actions of the ocean (spring tides, neap tides, and so on)? Because the relative positions of the earth, the moon, and the sun determine both the phases of the moon and the combined strength of the gravitational forces of attraction exerted on the ocean water by the moon and the sun. So the changes in gravitational force are the proximate causes of tidal actions, and the relative positions of the three bodies can be thought of as their distal cause. The phases of the moon are merely collateral effects of the positions of the three bodies involved and serve only as an indication of what the positions are (full moon when the earth is between the sun and the moon on a straight line, and so on), having no causal role whatever on tidal actions.

What about explaining, or interpreting, mind-brain correlations? Which of the models we have surveyed best fits the mind-body case? As we would expect, all of these models have been tried. We begin with some causal approaches to the mind-body relation:

**Causal Interactionism.** Descartes thought that causal interaction between the mind and the body occurred in the pineal gland (chapter 2). He speculated that “animal spirits”—fluids made up of extremely fine particles flowing around the pineal gland—cause it to move in various ways, and these motions of the gland in turn cause conscious states of the mind. Conversely, the mind could cause the gland to move in various ways, affecting the flow of the surrounding animal spirits. This in turn influenced the flow of these fluids to different parts of the body, ultimately issuing in various physiological changes and bodily movements.¹

“Preestablished Harmony” Between Mind and Body. Leibniz, like many of his great contemporary Rationalists, thought that no coherent sense could be made of Descartes’s idea that an immaterial mind could causally influence, or be influenced by, a material body like the pineal gland, managing to move this not-so-insignificant lump of tissue hither and thither. On his view, the mind and the body are in a “preestablished harmony,” rather like the clocks that were synchronized by the shopkeeper in the morning, with God having started off our minds and bodies in a harmonious relationship. Whether this is any less fantastical an idea, at least for us, than Descartes’s idea of mind-body interaction is debatable.
Occasionalism. According to Nicolas Malebranche, another major Continental Rationalist, whenever a mental event appears to cause a physical event or a physical event appears to cause a mental event, it is only an illusion. There is no direct causal relation between “finite minds” and bodies; when a mental event, say, your will to raise your arm, occurs, that only serves as an occasion for God to intervene and cause your arm to rise. Divine intervention is also responsible for the apparent causation of mental events by physical events: When your finger is cut, that again is an occasion for God to step in and cause you pain. The role of God, then, is rather like that of the leprechaun in the clock shop whose job is to keep the clocks synchronized at all times by continuous interventions. This view is known as occasionalism; it was an outcome of the doctrine, accepted by Malebranche and many others at the time, that God is the only genuine causal agent in this world, and that the apparent causal relations we observe in the created world are only that, an appearance.

The Double-Aspect Theory. Spinoza, another great Rationalist of the time, maintained that mind and body are simply two correlated aspects of a single underlying substance that is in itself neither mental nor material. This theory, like the doctrine of preestablished harmony and occasionalism, denies direct causal relationships between the mental and the physical; however, unlike them, it does not invoke God's causal action to explain the mental-physical correlations. The observed correlations are there because they are two distinguishable aspects of one underlying reality. A modern form of this approach is known as neutral monism, according to which the fundamental reality is neutral in the sense that it is intrinsically neither physical nor mental.

Epiphenomenalism. According to T. H. Huxley, a noted British biologist of the nineteenth century, all conscious events are caused by neural events in the brain, but they have no causal power of their own, being the ultimate end points of causal chains. So all mental events are effects of the physiological processes in the brain, but they are powerless to cause anything else—even other mental events. You “will” your arm to rise, and it rises. But to think that your volition is the cause of the rising of the arm is to commit the same error as thinking that the changes in the phases of the moon cause the changes in tidal motions. The real cause of the arm's rising is a certain neural event in your brain, and this event also causes your experience of a volition to raise the arm. This is like the case of the moon and the tides: The relative positions of the earth, the moon, and the sun are the true cause of both the tidal motions
and the phases of the moon. Many scientists in brain research seem to hold, at least implicitly, a view of this kind (see chapter 10).

Emergentism. There is another interesting response to the question “Why are mental phenomena correlated with neural phenomena in the way they are?” It is this: The question is unanswerable—the correlations are “brute facts” that we must simply accept; they are not subject to further explanation. This is the position of emergentism. It holds that when biological processes attain a certain level of organizational complexity, a wholly new type of phenomenon, namely, consciousness and rationality, “emerges,” and why and how these phenomena emerge is not explainable in terms of the lower-level physical-biological facts. There is no explanation of why, say, pains rather than itches emerge from C-fiber activations or why pains emerge from C-fiber activations rather than another type of neural state. That there are just these emergence relationships and not others must be accepted, in the words of Samuel Alexander, a leading theoretician of the emergence school, “with natural piety.” The phenomenon of emergence must be recognized as a fundamental fact about the natural world. One important difference between emergentism and epiphenomenalism is that the former, but not the latter, acknowledges causal power and efficacy of emergent mental phenomena.

The Psychoneural (or Psychophysical, Mind-Body) Identity Theory. This position, explicitly advanced as a solution to the mind-body problem in the late 1950s, advocates the identification of mental states with the physical processes in the brain. Just as there are no bolts of lightning over and above atmospheric electrical discharges, there are no mental events over and above, or in addition to, the neural processes in the brain. “Lightning” and “electrical discharge” are not dictionary synonyms, and the Greeks probably knew something about lightning but nothing about electric discharges; nonetheless, bolts of lightning are just electric discharges, and the two expressions “lightning” and “atmospheric electric discharge” refer to the same phenomenon. In the same way, the terms “pain” and “C-fiber activation” do not have the same dictionary meaning; Socrates knew a lot about pains but nothing about C-fiber stimulation. And yet pains turn out to be the activations of C-fibers, just as bolts of lightning turned out to be electrical discharges. In many ways, mind-brain identity seems like a natural position to take; it is not just that we point to our heads when we are asked where our minds are. Unless you are prepared to embrace Cartesian immaterial mental substances outside physical space, what
could your mind be if not your brain? And what could mental states be if not states of the brain?

* * *

But what are the arguments that support the identification of mental events with brain events? Even if your mind is in your head, your mind and your brain might only share the same space while remaining distinct. So are there good reasons for thinking that the mind is the brain? There are three principal arguments for the mind-brain identity theory. These are the simplicity argument, the explanatory argument, and the causal argument. We will see how these arguments can be formulated and defended, and try to assess their cogency. We will then turn to some arguments designed to refute, or at least discredit, the mind-brain identity theory.

THE ARGUMENT FROM SIMPLICITY

J. J. C. Smart, whose 1959 essay “Sensations and Brain Processes” had a critical role in establishing the psychoneural identity theory as a major position on the mind-body problem, emphasized the importance of simplicity as a ground for accepting the theory. He writes:

Why do I wish [to identify sensations with brain processes]? Mainly because of Occam’s razor. . . . There does seem to be, so far as science is concerned, nothing in the world but increasingly complex arrangements of physical constituents. All except for one place: in consciousness. That is, for a full description of what is going on in a man you would have to mention not only the physical processes in his tissues, glands, nervous system, and so forth, but also his states of consciousness: his visual, auditory, and tactual sensations, his aches and pains. That these should be correlated with brain processes does not help, for to say that they are correlated is to say that they are something “over and above.” . . . So sensations, states of consciousness, do seem to be the one sort of thing left outside the physicalist picture, and for various reasons I just cannot believe that this can be so. That everything be explicable in terms of physics . . . except the occurrence of sensations seems to me frankly unbelievable.

Occam’s (or Ockham’s) razor, named after the fourteenth-century philosopher William of Ockham, is a principle that urges simplicity as an important
virtue of theories and hypotheses. The following two formulations are among
the standard ways of stating this principle:

I. Entities must not be multiplied beyond necessity.
II. What can be done with fewer assumptions should not be done with
more.

Principle (I) urges us to adopt the simplest ontology possible, one that
posits no unnecessary entities—that is, entities that have no work to do. In
mathematics, we deal with natural numbers, rationals, and reals. But real num-
bers can be constructed out of rationals, which in turn can be constructed out
of natural numbers. Natural numbers, too, can be generated as a series of sets.
Sets are all we need to do mathematics. A crucial question in applying this
principle, of course, is to determine what counts as going “beyond necessity,”
or what “work” needs to be done. The physicalist would hold that Cartesian
immaterial minds are useless and unneeded posits; the Cartesian dualist, how-
ever, would disagree precisely on that point.

Principle (II) can be taken as urging simplicity and economy in theory
construction: Choose the theory that gives the simplest, most parsimonious
descriptions and explanations of the phenomena in its domain—that is, the
theory that does its work with the fewest independent hypotheses and as-
sumptions. When Napoleon asked the astronomer and mathematician Pierre
de Laplace why God was absent from his theory of the planetary system,
Laplace is reported to have replied, “Sir, I have no need of that hypothesis.” To
explain what needs to be explained (the stability of the planetary system, in
this instance), we do well enough with physical laws alone; we need no help,
and get none, from the “hypothesis” that God exists. Here, he is invoking ver-
sion (II) of Ockham’s razor. We can also see Laplace as invoking version (I):
We don’t need God in our ontology to do planetary astronomy; he would be
an idler with no work to do.

There seem to be three lines of consideration one might pursue in at-
tempts to argue in favor of the mind-brain identity theory on the ground of
simplicity.

First, it is a simple fact that identification reduces the number of putative
entities and thereby enhances ontological simplicity. When you say X is the
same thing as Y—or, as Smart puts it, that X is nothing “over and above” Y—
you are saying that there is just one thing here, not two. So if pain as a mental
kind is identified with its neural correlate, we simplify our ontology on two
levels: First, there is no mental kind, being in pain, in addition to C-fiber
stimulation; second—and this follows from the previous point—there are no individual pain occurrences in addition to occurrences of C-fiber stimulation. In this rather obvious way, mind-brain identification simplifies our ontology.

Second, it may also be argued that psychoneural identification is conducive to conceptual or linguistic simplicity as well. If all mental states are systematically identified with their neural correlates, there is a sense in which mentalistic language—language in which we speak of sensations, emotions, and thoughts—is in principle replaceable by a physical language in which we speak of neural processes. The mentalistic language is practically indispensable and we can be certain that it will remain so. We will almost certainly never have a full catalog of mental-neural correlations, and who among us will want to learn the bewilderingly complex and arcane medical terms? Still, we cannot deny the following crucial fact: On the identity theory, descriptions formulated in a mental vocabulary do not report facts or phenomena distinct from those reportable by sentences in a comprehensive physical-biological language. There are no excess facts beyond physical facts that can only be described in some nonphysical language. In this sense, physical language would be complete and universal.

Third, and this is what Smart seems to have in mind, suppose we stop short of identifying pain with C-fiber stimulation and stick with the correlation “Pain occurs if and only if (iff) Cfs occurs.” As earlier noted, correlations cry out for explanation. How might such correlations be explained? In science, we standardly explain laws and correlations by deriving them from other, more fundamental laws and correlations. From what more basic correlations could we derive “Pain occurs iff Cfs occurs”? It seems quite certain that it cannot be derived from purely physical-biological laws alone. The simple reason is that these laws do not even speak of pain; the term, or concept, “pain” does not appear in physical-biological laws, for the obvious reason that it is not part of the physical-biological language. So if the pain-Cfs correlation is to be explained, its explanatory premises (premises from which it is to be derived) will have to include at least one law correlating some mental phenomenon with a physical-biological phenomenon—that is, at least one psychoneural correlation. But this puts us back in square one: How do we explain this perhaps more fundamental mental-physical correlation?

The upshot is that we are likely to be stuck with the pain-Cfs correlation and countless other such psychoneural correlations, one for each distinct type of mental state. (Think about how many mental states there are or could be, and in particular, consider this: For each declarative sentence \( p \), such as “It will
snow tomorrow,” there is the belief that \( p \)—that is, the belief that it will snow tomorrow.) And all such correlations would have to be taken as “brute” basic laws of the world—“brute” in the sense that they are not further explainable and must be taken to be among the fundamental laws of our total theory of the world. (We will shortly discuss an argument, “explanatory argument I,” that claims that these psychoneural correlations are explained by psychoneural identities; for example, that “pain occurs iff Cfs occurs” is explained by “pain = Cfs.”)

But such a theory of the world should strike us as intolerably complex and bloated—the very antithesis of simplicity and elegance we strive for in science. For one thing, it includes a huge and motley crowd of psychoneural correlation laws—a potentially infinite number of them—among its basic laws. For another, each of these psychoneural laws is highly complex: Pain may be a “simple” sensory quality, but look at the physical side of the pain-Cfs correlation. Cfs consists of an untold number of molecules, atoms, and particles, and their interactions. We expect our basic laws to be reasonably simple, and reasonably few in number. And we expect to explain complex phenomena by combining and iteratively applying a few simple laws. We do not expect basic laws to deal in physical structures consisting of zillions of particles in unimaginably complex configurations. This makes our total theory messy, inflated, and inelegant.

Compare this bloated picture with what we get if we move from psychoneural correlations to psychoneural identities—from “pain occurs iff Cfs occurs” to “pain = Cfs.” Pain and Cfs are one and not two, and we are not faced by two distinct phenomena whose correlation needs to be explained. In this way, psychoneural identities permit us to transcend and renounce these would-be correlation laws—what Herbert Feigl aptly called “nomological danglers.” Moreover, as Smart emphasizes, the identification of the mental with the physical brings the mental within the purview of physical theory, and ultimately our basic physics constitutes a complete and comprehensive explanatory framework adequate for all aspects of the natural world. The resulting picture is far simpler and more elegant than the earlier picture in which any complete theory of the world must include all those complex mind-brain laws in addition to the basic laws of physics. Anyway, that is the argument.

What should we think of this argument? Does going from psychoneural correlations to psychoneural identities really simplify our total theory of the world, as the argument claims? Here the reader is invited to reflect on the following simple question: Doesn’t the psychoneural identity theory merely
replace psychoneural correlations with an equal number of psychoneural identities, one for one? The identities are empirical just like the correlations, and they make even stronger modal assertions about the world, going beyond the correlations. This is so because the identity “pain = Cfs” is now generally taken to be a necessary truth (if true), and the correlation “pain occurs iff Cfs occurs,” being entailed by a necessary truth, turns out itself to be a necessary truth. Moreover, these identities are not deducible from more basic physical-biological laws any more than the correlations are, and so they must be countenanced as fundamental and ineliminable postulates about how things are in the world. So don’t we end up with the same number of empirical assumptions about the world? The fact is that the total empirical content of a theory with psychoneural identities is at least equal to that of a theory with the psychoneural correlations they replace. Doesn’t it follow that version (II) of the simplicity principle actually argues against psychoneural identities, or declares a tie between the identities and the correlations? So what exactly are the vaunted benefits of simplification promised by the identities?

The reader is also invited to consider how a Cartesian, or a dualist of any stripe, might respond to Smart’s simplicity argument, keeping in mind that one person’s “simple” theory may well be another person’s “incomplete” or “truncated” theory. What counts as “going beyond necessity” can be a matter of dispute—in fact, what is to be included among “the necessities” is usually the very bone of contention between the disputants.

Explanatory Arguments for Psychoneural Identity

According to some philosophers, psychoneural identities can do important and indispensable explanatory work—that is, they help explain certain facts and phenomena that would otherwise remain unexplained, and this provides us with a sufficient warrant for their acceptance. Sometimes an appeal is made to the principle of “inference to the best explanation.” This principle is usually taken as an inductive rule of inference, and there is a widespread, if not universal, agreement that it is an important rule used in the sciences to evaluate the merits of theories and hypotheses. The rule can be stated something like this:

Principle of Inference to the Best Explanation. If hypothesis H gives the best explanation of phenomena in a given domain when compared with other rival hypotheses H₁, . . . , Hₙ, we may accept H as true, or at least we should prefer H over H₁, . . . , Hₙ.⁸