MIT Press • Open Encyclopedia of Cognitive Science

The Mind-Body Problem

Tim Crane[1]

[1]**Central European University**

MIT Press

Published on: Feb 10, 2025

DOI: https://doi.org/10.21428/e2759450.3614036a

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

The mind–body problem is the problem of explaining how our minds are related to our bodies and to the rest of the material world. It is obvious that our experiences and thoughts, feelings, and emotions are closely connected to our bodies and brains. But what exactly is the nature of this connection? Is your mind simply identical with your brain and central nervous system? *Physicalism* or *materialism* says yes. *Dualism* and *idealism* say no; dualism says that mind and matter are two fundamentally different kinds of thing, whereas idealism says that reality is ultimately mental in nature. The mind–body problem really consists of two subproblems. The first is the problem of causal interaction, primarily a problem for dualism: If mind and matter are such different things, then how can minds have effects in the material world? The second is the problem of consciousness, primarily a problem for physicalism: How can consciousness simply be something material, since it seems so unlike everything else in the physical world? Taken together, these two subproblems in effect form a dilemma: If the mind is not physical, then how is mental causation possible? But if the mind is physical, then how should consciousness be explained?

History

The main outlines of today's mind–body problem derive from René Descartes (1596-1650). Of course, there were important views of the relationship between mind (or soul) and body before Descartes. Perhaps most influential in the Western philosophical tradition was Aristotle (384-322 BC), who conceived of the relationship between the soul and body in terms of his theory of substance. A substance, for Aristotle, is a fundamental being; substances make up the ultimate reality of the world (Robinson & Weir, 2024). Every individual substance is a combination of matter (the "stuff" of which it is made) and form (that organizes it into a unified individual thing). For Aristotle, the soul is the form of the body's matter. In the interpretation of St. Thomas Aquinas (1225-1274), Aristotle's theory became the official philosophy of mind and body in the pre-Reformation Christian church.

In the 17th century, Descartes adopted a radically new conception of substance and, with it, a new conception of the mind–body relationship (Descartes, 1996). For Descartes, substance is defined simply as what is capable of existence independently of any other thing. Your mind, he believed, could exist independently of your body—therefore it is a separate substance. Your haircut, by contrast, cannot exist independently of your hair or your head—so it is not a substance but only a modification of your hair. Descartes believed that everyone's mind or soul is a substance and that the material world is a separate substance. This view, that the world consists of two totally distinct kinds of thing, is Descartes's dualism (known as *Cartesian dualism* after the Latin form of Descartes's surname, *Cartesius*).

Cartesian dualism was very influential but not universally accepted. It is rejected by *materialists*, who believe that the world consists only of matter, and there is no such thing as the Cartesian soul. Versions of materialism have been around since ancient Greek times. In the 17th century, materialists like Thomas Hobbes (1588-1669) and Pierre Gassendi (1592-1655) challenged Descartes's dualism, and the debate between materialists, dualists,

and idealists has continued until the present century. It would be quite wrong, then, to think that materialism about the mind is an invention of recent times as some scientists seem to think (Crick, 1994). Some philosophers in the early 20th century introduced the term *physicalism* to replace *materialism*, partly because physics recognizes the existence of many things that are not matter in the strict sense (e.g., spacetime, forces, and fields; Stoljar, 2024). These days, however, the terms are interchangeable in most contexts (Stoljar, 2024). From the mid-20th century onwards, physicalism became the dominant, yet not unchallenged, view (for some main alternatives, see Guyer & Horstmann, 2023 and Robinson, 2023).

Core concepts

Traditionally, the mind–body problem is expressed in terms of the question, "Are mind and body one thing or two?" But what is meant by "thing" here? As we saw, Cartesian dualists understood it in terms of the notion of substance, as did many materialists (see Robinson & Weir, 2024). But it can also be understood as concerning properties or features or attributes of things (for example, the property of being in pain or of experiencing thirst). This provides another way of thinking of dualism and physicalism. Dualisms can be divided into substance dualism and property dualism. Similarly, physicalism can be a view about substances—all substances are physical—or about properties too—all properties are physical. The view that all properties are physical is often known as *reductive physicalism*. However, this needs a little clarification, since the term "reductive" has been used in various different ways.

Reduction in the philosophy of science can either be an ontological or an epistemological idea. *Ontological reduction* involves the identification of entities at one level of scientific inquiry with entities at another level—as when scientists might identify the thermodynamic phenomenon of a gas's temperature with the mechanical phenomenon of its mean molecular kinetic energy. The gas's temperature is a phenomenon at the thermodynamic level, whereas its kinetic energy is at the mechanical level. Ontological reduction says that they are literally the same thing. *Explanatory reduction*, by contrast, involves the explanation of one theory by a theory at a lower level—in our example, this would be the explanation of thermodynamics in terms of statistical mechanics. So, ontological reduction is a relationship between entities, whereas explanatory reduction is a relationship between theories [see Mechanistic Explanation].

Given this understanding of reduction, we can say that reductive physicalism about the mind either identifies mental entities with biological or physical entities (this is *ontologically reductive physicalism*, or the *identity theory*), or it aims to explain psychology in terms of biology or physics (this is explanatorily reductive physicalism). Ontological reduction can involve identifying mental properties with physical properties (the *type identity theory*; Smart, 1959; Lewis, 1994; Armstrong, 1980) or identifying only mental objects/events with physical objects/events (the *token identity theory*; Davidson, 1970).

It is important to emphasize that ontological reduction in itself does not imply explanatory reduction. In other words, just because mental entities are physical entities, this does not mean that the mind can be explained in

terms of biology or biochemistry or physics. The claim is only that mental entities are identical with physical entities (see Davidson, 1970). So, it is possible to believe in the ontological reduction of the mental without believing in its explanatory reduction.

It is also possible to believe in the explanatory reduction of the mental without believing in ontological reduction, if you hold that mental entities are distinct from physical entities, but psychology can nonetheless be explained in terms of biology or physics. This would be an explanatorily reductive version of physicalism that is not ontologically reductive.

But what exactly is this ontologically nonreductive physicalism? It typically claims that although every substance—every object or every being capable of independent existence—is physical, not all properties are physical. A principal reason for denying that all properties are physical (i.e., denying the identity theory of properties, or the type identity theory) is the argument from variable or multiple realization (Putnam, 1975). It seems empirically unlikely that the same physical structures will underpin mental states across all organisms, but it seems very likely that at least some of these physically different organisms can instantiate the same mental properties and so will be in states of the same mental type. So, Putnam argued, we should reject the type identity version of physicalism on these empirical grounds. However, this is consistent with holding that all objects are physical.

Just as physicalism can be a thesis about properties or objects (substances), so too can we distinguish between substance dualism (Descartes's view) and property dualism, the view that there are two very different kinds of property in the world (Chalmers, 1996). Property dualism is a weaker form of dualism than substance dualism.

But since ontologically nonreductive physicalism denies the identity theory, is it not also a kind of property dualism? Not according to physicalism, which holds that everything, including the mental, supervenes on the physical (McLaughlin & Bennett, 2023). This supervenience is the idea that no two things can differ in their mental properties without differing in their physical properties ("no difference without a physical difference"). In other words, identical physical objects will also be mentally identical. Ontologically nonreductive physicalism is committed to this doctrine of supervenience: mental properties are distinct from physical properties, but they are deeply dependent on them. And since dualism in general requires ontological independence, supervenience therefore rules out property dualism.

Sometimes philosophers connect supervenience to the idea of *emergence*: mental properties "emerge" from the physical properties of the brain without being identical to those physical properties (McLaughlin, 1992; Vintiadis, 2013). Although the intuitive idea might seem clear enough, the term *emergence* is in fact understood in many different ways. Some thinkers understand emergent properties as any higher-level properties of a relatively complex system, which are not shared by all the system's parts (color might be emergent in this sense). Others use it to pick out a subset of those higher-level properties, which meet the further epistemological condition of not being predictable from knowledge of the lower level, despite being

ontologically dependent or supervenient on properties at that level. On this view, emergent properties depend on lower level properties as a “brute” or inexplicable fact (see Vintiadis & Mekios, 2018). Still others use *emergent property* for those higher-level mental properties that do not supervene with necessity on the lower level—in other words, properties as conceived of by property dualism (Chalmers, 1996). Partly because of this terminological confusion, talk of emergence has not always clarified the mind–body debate.

Questions, controversies, and new developments

The argument for physicalism: Mental causation

The main argument for physicalism—and against dualism—comes from the apparent fact of mental causation (Bennett, 2007; Elisabeth of Bohemia, 2017). It seems obvious that our minds (mental states, events, and processes) make things happen in the physical world—they make our bodies move. What we do is caused, in part, by what we want and how we think the world is. And yet, we seem to have learned from science that the physical world is causally closed: nothing can happen in the physical world that does not have a sufficient or adequate physical cause. It follows that if we trace the causes of physical effects back through history, we will not need to leave the realm of the physical. If this is so, then how is there any “room” for our minds to make things happen if they are (as dualism holds) distinct from our brains or bodies?

One possibility is that our minds make things happen by adding an additional cause to the physical cause. But then it would seem that mental and physical causes overdetermine their physical effects; for each of these effects, there would be two independent causes, each of which is enough to bring about the effect. This is not logically inconsistent with what we know. But it does seem an extraordinarily unlikely coincidence that whenever I decide to do something, my decision causes my action, and there are also independent mechanisms in my brain which make me do it. And yet, to embrace the *epiphenomenalist* idea that there is no mental causation at all—that our minds never make things literally change in the physical world—is even more unacceptable.

Physicalists therefore draw the conclusion that the mental and the physical causes of a physical effect are not distinct or independent but the very same thing (Montero & Papineau, 2015). Mental causes can make changes in the physical world, but this is because they are identical with physical causes in the brain. There are various forms of this argument, but this way of arguing from mental causation to the identity theory is the main philosophical argument for physicalism (Lewis, 1966; Davidson, 1970; Armstrong, 1980; Papineau, 2002). It is important to emphasize that to assert the identity of mental and physical entities is not to deny the existence of the mental (any more than asserting the identity of water and H_2O is to deny the existence of water). Ontological reduction is not the same as elimination.

The argument against physicalism: The problem of consciousness

The main challenge to physicalist theories of the mind comes from the phenomenon of consciousness (Chalmers, 1996; Crane, 2018). It may seem difficult to understand how unconscious thinking, representation, and reasoning are embodied in the brain. But consciousness seems even more deeply mysterious. What is responsible for us having experiences, sensations, emotions, and so on? How can mere brain activity be the basis of all this? It can seem that no matter how much we know about how the brain functions, we would still not know why conscious states feel the way they do. Or to use the famous illustration: no matter how much we knew about the anatomy and physiology of a bat's sensory system, we would arguably never know what it is like (what it *feels* like) to be a bat (Nagel, 1974).

This example is not intended to show that physicalism is false; rather, it is intended to show that even if it is true, physicalism is not really intelligible to us. Some philosophers have expressed this in terms of the idea of an *explanatory gap* (Levine, 1983). Generalizing from the bat example, the idea is that there is an inevitable epistemological gap between our knowledge of the brain and our knowledge of consciousness. However, many physicalists believe that the existence of such a gap does not threaten the truth of physicalism (Block & Stalnaker, 1999). What is needed, rather, is a physicalist explanation of why the epistemological or explanatory gap exists (Papineau, 2002).

However, there is another argument involving a thought experiment that is specifically designed to refute physicalism (Jackson, 1982). We are asked to consider a brilliant scientist called Mary, who knows all the facts about the physics and biology of color vision but who has never seen any colors apart from black and white. If one day Mary sees something red for the first time, then it seems that she learns something new: what red looks like. If what she has learned is a fact, then it seems that not all facts are physical facts, since by hypothesis she already knew all the physical facts about color. Therefore not all facts are physical facts, and physicalism is false.

The two arguments discussed above both rest on the possibility of detaching knowledge of the physical world (the bat's physiology and anatomy, the physical basis of color and color vision) from knowledge of consciousness (what it is like to be a bat, what it is like to see colors). There is a further, distinct argument that also appeals to the apparent possibility of detaching physical phenomena from consciousness as such (Chalmers, 1996). We are asked to conceive of a "zombie"—a physical replica of one of us but that lacks consciousness. Assume also that the conceivability of zombies implies their possibility. But if zombies are possible, then physicalism is false, since physicalism must at the very least imply that everything is determined, or necessitated, by how things are physically (this is the supervenience claim explained earlier). This necessitation means that reproducing something with the same physical nature must reproduce its entire nature. The mere possibility of zombies is a counterexample to this claim and therefore to physicalism.

The zombie argument relies on considerations about what we can conceive; for this reason, it is sometimes known as a *conceivability argument* (Balog, 1999). The main resistance to this argument involves skepticism about moving to conclusions about the ontology (or reality) of the mental from psychological assumptions about what we can conceive. Much of the dispute has therefore focused on the conceivability of zombies: whether they are really conceivable and, if so, whether this really does imply that they are possible (Papineau, 2002).

Some versions of these anti-physicalist arguments employ the idea that consciousness must involve the instantiation of irreducible, intrinsic, nonrepresentational (or even nonphysical) properties called *qualia*. This is confusing, however, because other philosophers have used qualia to refer simply to conscious properties themselves, however they are further theorized. So the term either expresses a controversial theory of consciousness (as Dennett, 1991 argues) or merely announces the existence of consciousness. This messy terminological situation can make it difficult to affirm the existence of consciousness without denying physicalism or to make sense of the views of those who reject the connotations of qualia while believing in consciousness (for example, those who think that consciousness is an aspect of mental representation).

The fact of the matter, however, is that none of the arguments presuppose that consciousness has the properties that the controversial theory of qualia says it has. All three of the arguments (concerning bats, Mary the color scientist, and zombies) can be adequately expressed without using the term qualia or the concept of qualia (see Crane, 2018).

Moving beyond physicalism and dualism

In the early 21st century, many philosophers tried to move beyond the apparent stalemate of the debate between physicalism and dualism. Some returned to an idea that emerged in the early 20th century, *neutral monism*, according to which being physical or mental is just a way of describing a reality that has an independent "neutral" nature (Russell, 1921; Chalmers, 1996). The reasoning here was based on the view that physics only describes the "structure and dynamics" of matter but not its intrinsic nature.

Panpsychism takes this argument in a different direction, agreeing with the view that physics can only describe the structure and dynamics of matter but arguing that its intrinsic nature must rather be some primitive form of consciousness (Goff et al., 2022). The basic idea is that the "qualitative" nature of consciousness cannot be captured by the "quantitative" theories of physics, yet consciousness must be embodied in matter somehow. *Panpsychists* conclude that consciousness is the intrinsic basis of all matter (Goff, 2019). The idea is controversial.

Some philosophers, by contrast, take the stalemate between physicalism and dualism to imply that the categories in which the debate is framed need to be reexamined. It is arguable that the exhaustive and exclusive mental/material dichotomy only really makes sense against the background of a 17th century conception of the

material world (Chomsky, 1995). Once we dispense with the idea that everything must have only one substantial nature (material or mental), then we open up the possibility of an ontological pluralism, which allows many kinds of entities and does not force a choice between physicalism, idealism, or dualism.

Broader connections

How does the mind–body problem connect with the systematic study of the mind in psychology, cognitive science, and neuroscience? Sometimes it is assumed that the empirical sciences of the mind must assume physicalism or ontological reduction. But this is not obviously the case; most of scientific psychology could work on the assumption that its psychological categories are perfectly in order without having to specify how they might reduce to the physical, in either the ontological or the explanatory senses [see Cognitive Ontology].

Of course, all cognitive scientists will believe that there are neural mechanisms underpinning our psychological capacities and will try and locate them using scientific methods. But this is consistent with being neutral on the materialism/dualism question or even with thinking that question is misconceived. If someone did take a property dualist view, they would probably have to commit to irreducible psychophysical laws (Chalmers, 1996), which is certainly an extra theoretical commitment. But in general, cognitive science does not need to take a stand on the mind–body question.

The same applies to new theoretical developments in cognitive science such as *predictive processing* or the *4E paradigms* (extended, embodied, enactive, and embedded cognition). Thinking of the mind in terms of any of the 4E concepts is compatible with ontologically reductive or nonreductive approaches. For example, the extended mind hypothesis could hold that the material over which the mind extends—the body and the environment—constitutes the supervenience basis of mental operations, or that its operations are identical with them, or that it just provides a mechanism for something which is genuinely distinct.

Similarly with predictive processing, this only specifies the overall architecture of mental operations—in essence, that mental processes function to minimize prediction error—but it does not specify whether the processes are ontologically reducible or whether they can be conceived within a property dualist framework.

Where the mind–body problem may seem to have empirical bite is in its application to scientific theories of consciousness. A number of scientific theories of consciousness take the challenge of explaining consciousness from a physicalist starting point; that is, since the entire world is physical, how should we explain the presence of consciousness in this world? The *integrated information theory* (IIT) of Tononi et al., (2016) attempts to answer this question by identifying a system's degree of consciousness with the level of integration of its constituent information. Another dominant theory of consciousness is Bernard Baars' *global workspace theory* (GWT; Baars, 1988), which treats consciousness in terms of the activity of representation in the brain's global workplace (see also Dennett, 1991).

GWT and IIT tend to focus on different locations in the brain, with GWT investigating the cognitive functions in the frontal areas, whereas IIT starts with the mechanisms of perception located largely in the back of the brain (Block, 2023). Despite these differences, both theories offer general explanations of the place of consciousness in the physical world and can therefore be seen as directly addressing the mind–body problem.

Acknowledgments

This work was funded in part by the Austrian Science Fund (FWF) [10.55776/COE3].

Further reading

- Chalmers, D. (1996). *The conscious mind.* Oxford University Press.
- Davidson, D. (1970). Mental events. In L. Foster & J. W. Swanson (Eds.), *Experience and theory (*pp. 79-101). Humanities Press.
- Descartes, R. (1996). *Meditations on first philosophy.* Cambridge University Press.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83*(4), 435-450. https://doi.org/10.2307/2183914

References

- Armstrong, D. M. (1980). The causal theory of the mind. In D. M. Armstrong (Ed.), *The nature of mind and other essays* (pp. 16-31). University of Queensland Press.

  ↩
- Baars, B. J. (1988). *A cognitive theory of consciousness.* Cambridge University Press.

  ↩
- Balog, K. (1999). Conceivability, possibility, and the mind-body problem. *Philosophical Review*, *108*(4), 497-528. https://doi.org/10.2307/2998286

  ↩
- Bennett, K. (2007). Mental causation. *Philosophy Compass*, *2*(2), 316-337. https://doi.org/10.1111/j.1747-9991.2007.00063.x

  ↩
- Block, N. (2023). *The border between seeing and thinking*. Oxford University Press.

  ↩
- Block, N. and Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review*, *108*(1), 1-46. https://doi.org/10.2307/2998259

  ↩

- Chalmers, D. (1996). *The conscious mind*. Oxford University Press.

  ↩
- Chomsky, N. (1995). Language and nature. *Mind*, *104*(413), 1-61. https://doi.org/10.1093/mind/104.413.1

  ↩
- Crane, T. (2018). A short history of philosophical theories of consciousness in the 20th century. In A. Kind (Ed.), *Philosophy of mind in the twentieth and twenty-first centuries: The history of the philosophy of mind* (vol. 6, pp. 78-103). Routledge.

  ↩
- Crick, F. (1994). *The astonishing hypothesis: The scientific search for the soul.* Scribners.

  ↩
- Davidson, D. (1970). Mental events. In L. Foster & J. W. Swanson (Eds.), *Experience and theory (*pp. 79-101). Humanities Press.

  ↩
- Dennett, D. C. (1991). *Consciousness explained.* Penguin Books.

  ↩
- Descartes, R. (1996). *Meditations on first philosophy.* Cambridge University Press.

  ↩
- Elisabeth of Bohemia (2017). Correspondence between Descartes and Princess Elisabeth. https://www.earlymoderntexts.com/assets/pdfs/descartes1643_1.pdf

  ↩
- Goff, P. (2019). *Galileo's error: Foundations for a new science of consciousness.* Pantheon Books.

  ↩
- Goff, P., Seager, W., & Allen-Hermanson, S. (2022). Panpsychism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2022 ed.). Stanford University.

  ↩
- Guyer, P., & Horstmann, R. F. (2023). Idealism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2023 ed.). Stanford University.

  ↩
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, *32*, 127-136. https://doi.org/10.2307/2960077

↩

- Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*(4), 354-361. https://doi.org/10.1111/j.1468-0114.1983.tb00207.x

  ↩

- Lewis, D. K. (1966). An argument for the identity theory. *Journal of Philosophy*, *63*(1), 17-25. https://doi.org/10.2307/2024524

  ↩

- Lewis, D. K. (1994). Reduction of mind. In S. D. Guttenplan (Ed.), *A companion to the philosophy of mind* (pp. 412-431). Blackwell.

  ↩

- McLaughlin, B. P. (1992). The rise and fall of British emergentism. In A. Beckermann, H. Flohr, & J. Kim (Eds.), *Emergence or reduction? Essays on the prospects of nonreductive physicalism* (pp. 49-93). De Gruyter.

  ↩

- McLaughlin, B., & Bennett, K. (2023). Supervenience. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Winter 2023 ed.). Stanford University.

  ↩

- Montero, B. G., & Papineau, D. (2015). Naturalism and physicalism. In K. J. Clark (Ed.), *The Blackwell companion to naturalism* (pp. 182–195). Wiley-Blackwell.

  ↩

- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*(4), 435-450. https://doi.org/10.2307/2183914

  ↩

- Papineau, D. (2002). *Thinking about consciousness.* Oxford University Press.

  ↩

- Putnam, H. (1975). The nature of mental states. In H. Putnam (Ed.), *Philosophical papers* (vol. 2, pp. 429-440). Cambridge University Press.

  ↩

- Robinson, H. (2023). Dualism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2023 ed.). Stanford University.

  ↩

- Robinson, H., & Weir, R. (2024). Substance. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Summer 2024 ed.). Stanford University.

  ↩
- Russell, B. (1921). *The analysis of mind*. George Allen and Unwin.

  ↩
- Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review*, *68*, 141-156. https://doi.org/10.2307/2182164

  ↩
- Stoljar, D. (2024). Physicalism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2024 ed.). Stanford University.

  ↩
- Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Review Neuroscience*, *17*, 450-461. https://doi.org/10.1038/nrn.2016.44

  ↩
- Vintiadis, E. (2013). Emergence. *Internet Encyclopedia of Philosophy.* University of Tennessee.

  ↩
- Vintiadis, E., & Mekios, C. (Eds.) (2018). *Brute facts.* Oxford University Press.

  ↩