

In Steven B. Cowan (ed.) *Problems in Epistemology and Metaphysics: An Introduction to Contemporary Debates*. New York: Bloomsbury Academic (2020), pp. 282—293.

### The Mind Is Material

Andrew Melnyk

University of Missouri

#### 1. What Is Materialism?

*Materialism* (or *physicalism*; the terms are interchangeable in philosophy) is the view that something you almost certainly believe to be true of your goldfish is also true of you and me, and of all human beings. The thing you almost certainly believe about your goldfish can be broken down into two claims: (i) a claim about your goldfish's *parts* and (ii) a claim about its *properties* (i.e., its attributes or characteristics) and the properties of its parts.

The *first* claim is that your goldfish is made up of certain bodily systems—its digestive system, its circulatory system, its muscular system, its nervous system, its immune system, and so on. Each of these systems is in turn made up of organs; for example, your goldfish's circulatory system is made up of its heart, gills, and blood vessels. Each of these organs is made up of specialized cells of about two hundred kinds, and each of these cells is made up of organelles and various molecules. The organelles are made up of molecules too, and all molecules are made up of atoms, which are made up of fundamental physical particles such as electrons and quarks. This first claim is not just that your goldfish has all these parts, but that it has no others; parts of the kinds mentioned are the only kinds of parts it has.

The *second* claim is that all the properties of your goldfish and of its parts (bodily systems, etc.) fall into one of two groups. In the first group are properties (e.g., the properties of electrons and atoms) that *physics* talks about—such properties as mass, charge, or spin. We may call such properties *physical*. In the second group are properties (e.g., the properties of cells and organs) that I'll call *functional* properties. For our purposes, let us say that a property of something (e.g., an organ) is a *functional* property if having the property is just a matter of its having parts that are *organized* in a certain way, e.g., organized so as to produce a particular *effect*. Your goldfish has the property of metabolizing glucose; but its having this property is just its having parts (e.g., insulin molecules, glucose transporters, insulin receptors) whose interaction has the net effect of converting glucose into energy that cells can use. Your goldfish's heart has the property of being a heart (obviously!); but its having this property is just its (cardiac) cells' being organized—related to one another—in such a way as to form an organ capable of drawing in and then pushing out blood.

Materialism is the view that what these two claims say about your goldfish is also true of every human being. So materialism is the view that:

(i) every human being can ultimately be broken down into parts all of which are physical, and

(ii) every property possessed by a human being, or by its parts, is either a physical property or a functional property in the sense already explained.

For the sake of a handy abbreviation, we can say that materialism is the view that every human being is *purely physical*.

Confusingly, the term, “materialism”, is also used in everyday life to refer to the view that the most important thing in life is acquiring wealth. But materialism in the sense of this essay doesn’t make any claim about how people *ought* to live their lives; it only makes a claim about how the world in fact *is*. So materialism in the sense of this essay is a different view from materialism in the everyday sense. Materialism in the sense of this essay might still lead *indirectly* to materialism in the everyday sense; but if it does, that would need to be shown by an argument.

The truth of materialism isn’t at all obvious; on the contrary, it initially strikes most people as crazy. Nevertheless, a case can be made for it. Making a case for materialism just means providing evidence for it—providing reasons to believe it. It doesn’t mean proving it, either in the mathematical sense in which Pythagoras’ Theorem can be proved by deducing it from self-evident axioms or in the weaker sense of providing so much evidence for a claim that there’s no room for reasonable doubt. Neither materialism nor its rivals can be proved in the mathematical sense. And if either materialism or its rivals could be supported by evidence that left no room for reasonable doubt, there wouldn’t be very smart people who accept materialism *and* very smart people who reject it; but of course there are both.

My case for materialism will be as follows. In section two, I clarify what does and doesn’t follow from the claim that we humans are purely physical. In sections three and four, I tackle the obvious objection that we cannot be purely physical because we have *minds*. Section

three spells out in a little detail what a materialist view of the mind would look like, while Section four offers some positive evidence for thinking that our minds are purely physical.

## 2. Two Common Reactions

One common reaction to materialism is to think, “But if materialism is true, we’re all just *material objects*, like rocks or tables”.

The trouble with this reaction is that rocks and tables aren’t typical of all material objects. They have rather simple physical constitutions, and, of course, they don’t really *do* anything. But not all material objects are like that. Your digestive system is a material object, but it can break down and process a huge variety of different foods and turn them all into energy and molecules that your body can use. A plane’s autopilot is a material object, but it can fly and land an airliner. The arrangement of computers and servers that stands behind Apple’s “Siri” virtual assistant is a material object, but it can do jobs that until recently would have required a human assistant. Unlike rocks or tables, these material objects have highly complex physical constitutions, which enable them to react to their respective environments (e.g., different foods, different altitudes, different commands) with elaborate and appropriate behavior. If humans *are* material objects, they’re material objects with a unique physical nature of immense complexity and a unique suite of abilities.

On the internet, I’ve seen a similar argument: if humans were material objects, then material objects like bottles of soda would have minds, just as humans do. But this argument misunderstands what materialism is committed to. Materialism says that all humans are material objects, not that all material objects are humans. And the claim that all humans are

material objects doesn't *entail* that all material objects are humans—just as the claim that all cats are mammals doesn't entail that all mammals are cats. So materialists aren't forced to say that all material objects are humans—so they aren't committed to the crazy claim that bottles of soda have minds.

A second common reaction to materialism is to think, “But if materialism is true, we're just *machines*”. But why couldn't we be machines? People feel sure that humans couldn't be machines because they've noticed two significant features of human behavior.

The first is that, when a human receives the *same* stimulus on two *different* occasions, he or she doesn't necessarily respond with the same behavior on both occasions. For example, on one occasion you might respond to, “Would you like a cookie?”, by saying, “Yes, please”, and taking a cookie, and then, five minutes later, respond to, “Would you like a cookie?”, by saying, “No, thanks; I just had one” and not taking a cookie. Or on one occasion you might respond to, “Who is the current prime minister of the UK?”, by saying, “Sorry, I don't remember”, and then respond to the very same question five minutes later, having just read a newspaper, by saying, “Theresa May”. Machines, it seems, don't act like this: they respond to the same stimulus with the same behavior. For example, whenever a doorbell's button is pressed, it rings.

The second feature of human behavior is that a human can receive *different* stimuli on two occasions and yet on both occasions respond in the *same* way. For example, you could learn that you'd won a scholarship either by reading an email or by hearing a voice on the phone. The physical characteristics of the two messages would be very different, the first being a pattern of light on a screen, the second a (different) pattern of vibrations in the air. But you

would react in the same way, by calling your parents, say. A machine, it seems, couldn't appreciate that the two stimuli were in some sense the same and merited the same behavioral response.

So do these two features of human behavior show that humans couldn't be machines? They don't. The two features of human behavior are perfectly genuine, but the assumption that the behavior of machines couldn't possibly have those features is false. Let us consider each feature in turn.

Feature #1. It's not true that machines always respond to the same stimulus with the same behavior. Many machines do, of course, like the doorbell, but many don't. An example of one that doesn't is your laptop, when it's running *Word*, and connected to a printer. On different occasions when you click on "Print", the laptop behaves differently, sending different signals to the printer. The signal it sends depends both on earlier stimuli—what keys were pressed when the document was typed—but also on the current internal state of the machine—what application it's running. Another example of a machine that can respond to the same stimulus with different behavior is a one-armed bandit at a casino. Nearly always, when you put in the money and pull the handle, you win nothing; but occasionally, when you put in the money and pull the handle, you win the jackpot.

In general, what behavior a machine produces depends on two things: (i) the stimulus the machine is currently receiving and (ii) the internal state of the machine. And the internal state of the machine is itself determined by two factors: (1) all the *past* stimuli the machine has received and (2) the internal state of the machine before it received its first stimulus.

We can now see why the doorbell can only respond with a single behavior when its button is pressed. The doorbell is sensitive to just *one* stimulus: the pressing of its button. And its internal state never changes. True, you change its internal state when you press its button, but only for as long as you hold down the button. You don't make a *lasting* change to the internal state of the doorbell that would make a difference to what behavior the doorbell produced the *next* time its button was pressed.

If humans are machines, then our behavior at a particular time depends on (i) the multiple influences that our environments have on our eyes and ears (and other sense organs) at that time and (ii) the internal state of our brains at that time (since human behavior depends in the first instance on what happens in our brains). And the internal state of our brains at a particular time depends on (1) the history of environmental influences on our brains up to that time, interacting with (2) the gene-guided development of our brains from before birth up to that time. One consequence: even if humans are machines, no two people's brains are exactly the same in both these respects, so we shouldn't expect that two people will always, or even often, react in exactly the same way to the same stimulus. Another consequence: even if humans are machines, we shouldn't expect to be able to *predict* a human's behavior in detail. To do so, we would need to know the exact state of every neuron in his or her brain, plus how the brain will be affected by every feature of his or her current environment. Obviously we have no way of knowing such things.

Feature #2. Let's now turn to the assumption that a machine couldn't react with the same behavior to two different stimuli—for example, to good news communicated in speech and the same good news communicated in writing. Again, many machines are not like that.

One example is an Android phone on which the Google search function can be controlled either by speaking words or by typing them; Google will perform the same search whether the instruction takes the form of sounds or of screen touches. (It doesn't matter whether the phone *understands* the spoken or typed words. The point is that the phone can respond to dramatically different stimuli with the same appropriate behavior.) Consider also the capacity of a dog to learn that its owner has come home by hearing a key in the front door or by seeing the owner walk toward the house or perhaps by catching the owner's scent from a distance; these are three very different stimuli eliciting the same behavior. If you think that dogs are purely physical (like goldfish), then dogs are also examples of machines that can react with the same behavior to very different stimuli.

### 3. What About Our Minds?

We have minds. That is, we *believe* that boarding starts in ten minutes, we *hope* that we have packed the right clothes, we *fear* that the pipes might leak while we are away, we *want* to have a safe flight, we *expect* that the flight will be uneventful. These mental states are called by philosophers *propositional attitudes*, because they consist of an *attitude* (e.g., belief, hope, fear, wanting, expectation) toward a *proposition* (e.g., that boarding starts in ten minutes, that we have packed the right clothes, or that the flight will be uneventful). But there is more to having a mind than propositional attitudes. We also dream and hallucinate; we have yellow afterimages after looking at a bright light; we feel pain, nausea, tingling, dizziness; we smell the smell of gasoline, taste the taste of mustard, see the redness of ketchup, feel the smoothness of glass. These mental states are called by philosophers *sensations*, because they seem to be

episodes of sensing or feeling. And of course we also *reason*, that is, go through thought processes. We can think to ourselves that no one at all is in the classroom at 11.05 a.m., that the professor said something at the start of semester about canceling a class, and that therefore class is probably cancelled today; in this example, we are reasoning to a *factual* conclusion, about what *is* the case. We can also think to ourselves that the needle of the gas gauge is in the red zone, and that therefore we should buy some gas at the next exit; in this example, we are reasoning to a *normative* conclusion, about what *should* be the case.

Now if materialism is true, human minds and human mental states must be purely physical. How could that could possibly be? I will try to explain, starting with *mental states*, and then turning to *minds*.

Let me start with *believing*, by which philosophers just mean *thinking that something is the case* (e.g., thinking that it's sunny). Our behavior (e.g., going to the store) can be explained in part by what we believe to be the case about our environments (e.g., that we're out of milk). So it's natural to think of *believing* something as having inside one's head "a map...by which we steer".<sup>1</sup> Think of how having a map of the campus helps you get around the campus. Beliefs are *map-like* in the sense that they represent the world (our bodies as well as our environments) as being a certain way, whether accurately or inaccurately. And we *steer* by beliefs because they play a part in guiding our behavior. But it looks like a purely physical state of the brain—the activation of a particular neural circuit, say—could play these two roles. A brain state could count as representing a certain state of affairs because, under the right conditions, it is caused by that state of affairs. And obviously brain states guide our behavior, by sending electrical signals along motor neurons to the right muscles at the right times.

*Desiring* can be viewed as a different kind of map by which we steer. The difference is that desires represent not a currently actual state of affairs but the state of affairs *desired*. For example, your desire for a soda represents your having a soda. And a desire guides behavior by having the job of helping to bring about the state of affairs represented. For example, your desire for a soda has the job of helping to bring it about that you have a soda.

*Sensations*, too, can be thought of as a kind of mental representation. A perceptual sensation (e.g., a visual sensation of red) can be regarded as a representation of our immediate environment—of things *outside* our bodies *right now* (e.g., a red object). A bodily sensation (e.g., a pain in the left foot) can be regarded as a representation of things *inside* our bodies right now (e.g., potential or actual tissue damage in the left foot).

Finally, *thought processes* can be thought of as *computations*—or at least as analogous to computations. What computers do is to take a string of complex symbols as input, perform operations on the symbols in accordance with a set of rules (a program), and then emit a different string of complex symbols as output. We're used to thinking of the symbols as 1's and 0's printed on a tape, or as magnetized and non-magnetized particles in a hard drive; but the symbols could equally well be different levels of electrical activity in neural circuits, and they will be if thought processes are computations.

Now for minds. The commonest materialist view of minds is that, strictly speaking, there are no such things—whether physical *or* non-physical! The idea is that we talk in everyday life about the human mind, and there's nothing wrong with that, but when we do, we're just talking about the *mental capacities* that we have—capacities to believe, to hope, to feel pain, and so on—rather than about an object. Aristotle, in his *De Anima*, was the first to

suggest such a view: he recommended replacing talk of humans' "having a soul" with talk of our "being ensouled", thus avoiding commitment to souls as objects. Two analogies may make the view more plausible. We may talk about Ruth's *personality*, but no one thinks that Ruth's personality is an *object* inside her, not even an *immaterial* object. We all understand that talk about Ruth's personality is just talk about her *personality traits*—traits like being friendly or cheerful or reserved or difficult. We may also talk about Ruth's *gait*, but we all know that a surgeon shouldn't expect to find Ruth's gait by opening up her body and looking inside. We know we're just talking about how Ruth walks—about the features that her walking has. Likewise for our talk of Ruth's mind: her mind is not an object inside her.

That's the commonest materialist view of minds. But I prefer a different view. On my view, people's minds are just as real as their circulations, or digestions, or immune systems. Indeed, we could call minds "*mental systems*", on the model of "*circulatory systems*", "*digestive systems*", "*immune systems*", and so on. And, just as your digestive system, say, is made up of an inter-connected series of organs—your stomach, liver, small intestine, and so forth—so also your mental system is (mainly) made up of your *brain*, itself an immensely complex network of sub-networks of brain cells constantly getting signals from our sense organs and sending signals to the body's muscles and to one another.

The digestive system has the function of converting food into usable energy and ingredients for bodily growth, and the immune system has the function of destroying harmful microbes. So what is the function of the mental system? The defining function of the human mental system, I suggest, has five components:

1. To use sensors to form descriptive representations of both the current external environment and the current internal states of the containing organism.

Examples of *sensors* would be the familiar five senses. *Descriptive representations* represent something as in fact being a certain way, i.e., as actually having a certain property. An example of a descriptive representation of the organism's current *external environment* would be a sensation representing—a sensation *of*—a red, round thing in front of you, or a belief that a tomato is in front of you. An example of a descriptive representation of the organism's current *internal states* would be an itch (which represents histamine in a part of your body) or a hunger pang (which represents that your stomach is contracting).

2. To undergo internal processes in which these sensory descriptive representations give rise to *further* descriptive representations of the organism's external environment and its internal states.

Examples of *internal processes* would be the rule-governed manipulation of descriptive representations to accomplish deductive or inductive reasoning. The *further descriptive representations* would be the conclusions of such reasoning, e.g., the belief that all tomatoes are red, or the belief that if a tomato is red, then something is red, or the belief that I need food.

3. To store some of these descriptive representations for later use.

The *storage* of representations would be something like *memory*.

4. To undergo internal processes in which these descriptive representations interact with one another and with representations of goal-states so as to produce organismic behavior, which behavior, often enough, achieves the organism's goals.

Examples of *internal processes* in this case would be the rule-governed manipulation of representations to accomplish *practical* reasoning, the kind of reasoning that supports a conclusion about what *ought* to be done. Examples of *representations of goal-states* would be desires; a desire for lunch, for example, would be a representation of your goal: having lunch. The representations manipulated in practical reasoning include both descriptive representations *and* representations of goal-states, reflecting the fact that, for example, we decide to walk to the freezer because we both *desire* ice cream and *believe* that there is ice cream in the freezer. The desire *alone* wouldn't make us decide to walk to the freezer; we might instead decide to drive to the store. And the belief *alone* wouldn't make us decide to walk to the freezer; we might be so full of pizza that we have no desire for ice cream.

5. To monitor the sensing and internal processing we engage in, and the descriptive representations and goal-state representations that we form, so as to form descriptive representations *of* these activities and states.

The idea here is that part of the defining function of a human mental system is to keep tabs on itself—on its own activities and states. Examples of “descriptive representations of these activities and states” would therefore be beliefs *about* one’s own sensations and beliefs and desires and thought processes.

The five sub-functions I have just distinguished can be thought of as working together to achieve a single overarching goal: that of gathering information about our world as it changes on timescales both short and long, and then using that information to generate behavior that succeeds in getting us what we want, whatever our world is like.

#### 4. Evidence That The Mind Is Material

I promised some positive evidence that our minds are purely physical. One kind of such evidence is the remarkable way in which mental states of every kind are *dependent* on purely physical states of the brain, namely, patterns of electrical activity in particular neural circuits. I don’t just mean the obvious point that, if our brains stopped working, our minds would stop working too. After all, if our *hearts* stopped working, our minds would stop working too, but that’s no evidence that mental states are nothing over and above *cardiac* states! I mean instead that, as far as we know, being in a particular mental state (whatever the mental state) requires—never occurs without—being in a particular brain state. We have learned this from a huge number of imaging studies. In such studies, the brains of experimental subjects are monitored with an imaging device (e.g., an fMRI machine) while the subjects are in certain mental states; the subjects are typically asked to perform a particular mental task, such as mentally rehearsing a learned motor skill, doing mental arithmetic, attending to an

unstimulated body part, and visualizing a scene. What is found in these imaging studies is that, whenever someone is in a particular mental state, there is always a particular brain state that the person is in at the same time. This sort of highly detailed dependence of mental states on neural states is exactly what we would expect to find if the mind is purely physical.

But it's *not* what we would expect to find if the most intuitively appealing non-materialist view—*interactionist substance dualism*—were true. It'll take a while to explain why. First, what is interactionist substance dualism? It's the view that your mental states are states of an *immaterial* entity—your mind—assumed to be distinct from your brain. So, for instance, when you're recalling childhood memories of your sister, that's because your immaterial mind is in a certain immaterial state; but if you start recalling childhood memories of your brother instead, that's because your immaterial mind has shifted into a different immaterial state. On this view, however, your mind still interacts causally with your brain, rather as an airline pilot causally interacts with the cockpit's instruments and controls, though distinct from them. So, for instance, when you notice an orange, the orange causes a change in the state of your visual cortex, which then causes a change in the state of your immaterial mind, and you have an experience of orange. Similarly, but in the opposite direction, when you decide to grasp a cup, that change in the state of your immaterial mind causes a change in the state of your motor cortex, which in turn causes muscles in your arm to contract.

Here's what's crucial: on this dualist view our brains serve *merely* to pass signals between our bodies and our immaterial minds. When we feel the heat of a fire, our brains relay signals *to* our immaterial minds *from* sensory neurons originating in our fingers. When we decide to step back from the fire, our brains relay signals *from* our immaterial minds *to* the

motor neurons that run to muscles in our legs. According to this dualist view, then, our brains *merely* play these mediating roles: in perception and in deciding to act. Our brains aren't where the *thinking* happens; our *immaterial minds* are. So if you're daydreaming, and you stop recalling childhood memories of your sister, and start recalling childhood memories of your brother instead, there's every reason on this dualist view to expect simultaneous change in the state of your *immaterial mind*, but no particular reason to expect any simultaneous change in the state of your *brain*—because daydreaming is neither perceiving nor deciding to act. The same goes for mentally rehearsing a learned motor skill, or doing mental arithmetic, or attending to an unstimulated body part, or visualizing a scene. Interactionist substance dualism doesn't predict that these mental activities will be accompanied by a distinctive kind of brain activity.

Interactionist substance dualism isn't the only kind of dualism. There is another kind—*property dualism*—and it *does* lead us to expect that mental states depend on neural states in the detailed ways we observe. Property dualism doesn't say that our mental states are states of an immaterial entity distinct from the brain; it says that mental states are states of the brain itself, but *immaterial* states of the brain—in other words, that the brain has certain *immaterial* properties in addition to its physical properties, these immaterial properties being our mental properties. Property dualism also says that the brain's immaterial mental properties march in lockstep with its physical properties; it posits laws of nature that bind each of the brain's immaterial mental properties to one of its neural properties in such a way that you can be in a particular immaterial mental state if, but only if, you're in a certain neural state; for example, you can feel cheerful if, but only if, a certain neural circuit in your brain is active. Because it

posits laws of this kind, property dualism leads us to expect that mental states depend on neural states in the detailed ways we observe.

So property dualism (unlike interactionist substance dualism) at least predicts the detailed ways in which mental states depend on neural states. But it's still *less plausible* than materialism in light of this dependence, because it can't explain the dependence as *economically* as materialism can—and the more economical of two theories that can explain the same facts is preferable. Property dualism can't explain the dependence as economically as materialism because property dualism must say that (1) immaterial properties of brains exist *in addition to* the physical properties of brains that everyone agrees exist, and that (2) laws of nature hold (binding the brain's immaterial to its physical properties) *in addition to* the laws of nature that everyone agrees hold.

There is other evidence that our minds are purely physical, which I can't go into here, but it also takes the form of facts that can be explained more economically by materialism than by any kind of dualism.

3-15-17

---

<sup>1</sup> As the philosopher and mathematician Frank Ramsey put it in his 1929 paper, "General Propositions and Causality", reprinted in *F. P. Ramsey: Philosophical Papers*, ed. D. H. Mellor (New York: Cambridge University Press, 1990).