What it is to have a mind: the mind-body problem from Descartes to a completed future physics

Christofer BULLSMITH

Consider how you can think of yourself in two different ways. On the one hand, you can think of yourself as a body in space and time, apt to displace water in the bathtub and benefitting from airbags in the event of an automobile accident. On the other hand, you can think of yourself as a locus of mental life (emotions, sensations, thoughts, and consciousness). The intuition that these two ways of thinking of yourself are incommensurable introduces a kind of dualism. On this view, my consciousness, subjective and immediate, could never be a mere physical process occurring in my body; my feeling of pain, or existential ennui, could surely not be seen, or comprehended in the rich and immediate way I do, by a third party with a scalpel and a microscope. We can cultivate this dualistic intuition by considering the seeming possibility of a consciousness which exists (in some sense) in the absence of any body.

In response, Descartes (1641) posited two kinds of substance, non-physical mind and material body, while Leibniz (1714: §17) posited mental and physical realms. This differentiation is in various ways enshrined in our language and ways of thinking, but dualism faces the challenge of explaining how mind and material (spirit and flesh, soul and body) interact. That is, we must give some account of how when the body is touched, the mind registers a sensation; of how when the mind wills an action, the body moves. Princess Elisabeth of the Palatinate wrote to Descartes in 1643 (see for example Shapiro 2008) and argued the need for such an account, and a satisfactory answer is seemingly yet to be given. This tension between the dualistic intuition that mind is somehow completely different from base matter, and the need to give an account of how this special mind interacts with base matter, is the root of the mind-body problem.

The natural way to characterize the problematic interactions between physical and non-physical is *causal*: the touch causes the sensation, and the will causes the movement. However, being both affected by physical causes and a cause of physical effects would seem to be a reasonable definition of what it is to be physical. If so, the mind is either physical and causally potent, or non-physical and causally impotent. Put another way, if physical effects always have physical causes, then any non-physical events are causally impotent in the physical world. A non-physical mind that is unaffected by the physical world and has no effect on the physical world seems to have no explanatory or predictive value ... in which case we'd presumably be better off not positing it in the first place.

Indeed, as our understanding of the information processing functions of nervous systems (of cognitive science and neuroscience, of the physical causal chains involved in cognition) improves, there seems to be ever less need or space for a mysterious non-physical mind stuff. If, for example, we can gain a complete understanding of the behaviour of a simple organism (say, the *C. elegans* roundworm being simulated at OpenWorm.org) using a bottom-up comprehensive computational simulation of the causal system consisting of its cells and immediate environment, then we have a purely physical causal story of how complex behaviour can arise without mind stuff. With roughly 1000 cells the millimetre-long *C. elegans* solves basic problems to crawl and swim about feeding, finding mates, and avoiding predators. More complex animals with more cells can display more complex behaviours, but our understanding of evolutionary biology leads us to expect the same kinds of physical causal systems: more cells, more connections. While there may be interesting discontinuities or watersheds along our phylogenetic branch of evolutionary history, or during individual ontogeny – becoming multicellular, gaining the first neuron or certain formats of intercellular connection that enable new forms of processing, and so on – there is simply no need or space for the sudden appearance of a new mental substance.

Now, all this discussion of neurons, computational simulation, and evolutionary biology prompts an easy response to the confusion over what the mind is: perhaps it's just the brain. On this view, if the mind is not some spooky non-physical stuff with no explanatory or predictive value, then it's physical stuff that is apt for causal, scientific study. We know that the mind and the brain are closely connected: changes to the brain, for example through growth or injury or the application of chemicals, can be intimately and systematically mind-altering. On this view, while the traditional concepts of mind and brain may be distinct (we wouldn't talk of a mind-surgeon, or of eating lamb's mind for dinner), we have now discovered that the mind is in fact the brain.

However, there are a number of reasons to think that minds are not simply brains.

1. Minds include non-brain bits of the body, and don't include some bits of the brain

Firstly, while the mind and the brain are clearly closely connected, and this connection motivates us to identify minds as brains, there doesn't seem to be any principled way of making out which parts of the nervous system exactly we ought to include. Presumably neurons in the brain are part of the mind, because they are directly involved in the relevant activities (input-output cognitive processing). Perhaps the bones of the skull are not part of the mind, playing at most an indirect support role … but these relevant mindful activities would not occur if the bones were removed, since brains not fixed in formalin have something like the consistency of cold porridge. Likewise, perhaps the blood vessels are not part of the mind, since they play a support role … but all activity would cease quickly without them. Any account of the mind-as-brain needs to give a principled account of what structures are in the mind, and which not, and why.

Further, various parts of the body which are not part of the brain have a *prima facie* claim to be included in the mind. An obvious candidate is the entire spinal column, which sits inside what is essentially the same protective membrane as the brain (the meninges), includes about one hundred million neurons, and processes information in much the same way that the brain does. Good candidates include the eyes, retinas, and the optical nerves, as these are intimately involved in early visual processing. Less obvious but still credible candidates include muscles (for example, people who frown during cognitive activity try harder and experience more cognitive strain, and forcing a smile can improve mood, seemingly indicating that the muscles of the face are involved in cognition; see for example Kahneman 2011, chapter 12). Indeed, even if neurons are privileged in some way that precludes muscles and blood vessels being considered part of the mind, the enteric nervous system (running through the gastrointestinal tract) contains about five hundred million neurons, about as many as the whole brain of a marmoset. A good final example might be the chemicals influencing the physiology of the brain and firing of neurons: the various structures around the body creating and regulating these chemicals ultimately play an important role in cognition too.

There's no great mystery here. If we use a digital computing analogy for the mind, the brain corresponds to the CPU. But this is a CPU which has spent several million years co-evolving with all the input, output, and memory systems of a specific chassis (a social, sexual chassis), such that the successful operation of the CPU depends intimately on specific characteristics of those co-evolved systems, and we can be genuinely unsure as to where the CPU ends and input/output/memory systems begin; or, put another way, the non-CPU components play critical roles structuring inputs and outputs, such that any story about how information processing occurs needs to include these non-CPU components. Indeed, if we use instead a more biologically appropriate connectionist (parallel distributed processor or neural network) computer analogy for the mind, there is no CPU. Instead, processing occurs as patterns of activation spread across layers of nodes, and a characterization of the system needs to include all levels from input (largely sensory receptors) to output, without any necessary privileging of the levels which happen to be inside the skull.

My own opinion here is that as our appreciation for the complex interactions involved in the development and operation of the human body develops, we will feel less and less able to draw the sharp distinction between 'important for cognitive processing' and 'not important for cognitive processing' which the mind=brain claim requires. The central nervous system outruns the brain; the peripheral nervous system outruns the central nervous system; and both are tied into complex physical and chemical systems extending around the body without which they would not function. If there is any principled way of making out what should be in or out with regard to 'important for cognitive processing', it may be at the boundary of the organism as a whole. That is, minds may be something that people have, not that brains are.

2. Minds include bits of the world outside not only the brain but also the body

One sense in which minds could include bits of the world might be if minds are essentially relational. There need be nothing spookily dualistic about relational concepts: for example, a 'brother' is a brother not by virtue of its own intrinsic properties (for example, being male), but also in virtue of standing in a particular relation to another thing (a sibling). Siegel (2016), for example, argues that mind is relational: that mind involves a sense of self in contrast to and as mirroring other people, and a flow of information between the self and others. Though interesting, Siegel's arguments seem to mostly be that putting the social nature of minds central stage has psychological and clinical returns (for example, encouraging empathy over competition, and belonging over solitude). Still, at the very least, Siegel's position serves to remind us that making sense of minds requires us to refer to the world, and especially to other people and social or cultural constructs such as language.

Unsurprisingly, philosophers have made similar points. Wittgenstein (1953: §243-315) argued that the idea of a private language is inconsistent: that language serves a social function and is tied to public (social, shared) criteria of use. With mental terms tied to behavioural criteria, even something as superficially simple as 'I think ...' can only be understood by bringing a rich social context of interpersonal identity and behavioural – meaning *physical* – criteria into view. As for Siegel, for Wittgenstein the mind is embodied and richly social.

Putnam (1973) and Burge (1979) argued for a position called semantic externalism – basically, that meanings are not entirely determined by something in the head, since they can change with contingent facts about the world (for Putnam, the microstructure of natural kinds; for Burge, the beliefs of others in my community) even as everything in the head is held constant. If semantic externalism is correct – and the broad consensus amongst philosophers seems to be that it is – then making out mental content requires reference to the world outside the head.

My own opinion here is that getting any *explanatory* handle on the mind requires reference to rich social and linguistic facts (not to mention historical facts – given that minds develop over time, and have evolved in some historical environment, even making out the basic fact that something is a fear rather than a hope seems to require bringing into view how fears and hopes have properly functioned historically in ancestral minds). However, I am not convinced that explanation requiring reference to the world constitutes minds 'including' these parts of the world: after all, understanding the kidney arguably requires reference to the organism as a whole (a kidney which does not serve to filter the blood of an organism is no kidney), but the kidney does not thereby 'include' the whole organism. Still, a kidney without an organism is not a (true, functioning) kidney, and in a similar way, a mind without a society is not a (true, functioning) mind.

There may be another simpler way in which minds could include bits of the world (other than brains and people): if minds include aids to cognition, for example to memory, to calculation, and to visualization. On this view, when I use a pencil and paper to diagram a problem in geometry, my scribblings become necessary to giving a full and comprehensible account of my cognitive activity, and therefore are to be considered a temporary part of my mind. Books and search engines can presumably do likewise; and even more interestingly, if pencil and paper or books can be made out as parts of my mind, then surely other people can too, for I can gain information, bounce questions, chant oral history contrapuntally, and generally engage synergistically with other people. And if other individuals can be intermittently parts of my mind, then presumably intermittent parts of their mind become parts of mine too, as my expert asks her expert, and so on.

There is, I think, something to be said for this view. Scribblings on paper might seem the wrong kind of thing to be part of my mind, since I don't have direct introspective access to them. But brain activity is not as unified and coherent as our subjective experience might suggest: Freud suggested in the 1900s that 90% of the mind was non-conscious, and subsequent neurological studies show the brain as a mess of competing interpretations and calls for action which riot continuously, with only the winner of the moment becoming available to conscious introspection. I don't have direct introspective access to these non-conscious subsystems of my mind, and I am not necessarily aware of where or even when they occur even should some output win its way to consciousness. Since I don't have direct introspective access to most of my mind, not having direct introspective access to the entirety of a process (because some of it is on paper, or involves another person) cannot disqualify it from being part of my mind.

The idea that minds can include bits of the world also, I think, helps future-proof our concept of mind. It is easy enough to imagine some small artificial pseudo-neuronal structure added to the brain, perhaps to return function after damage. As something in the head and neuron-like, and even functionally identical to the old neurons, we would presumably want to allow that this is part of the mind. But then it is surely easy to extend our imagining to structures further and further removed, in location or similarity to neurons. For example, technical constraints might mean we need to add the prosthesis inside the skull but displacing some other structure; or that we use transmitters inside the damaged brain and locate the artificial neurons just outside the skull but under the skin; or attached to the outside of the skin; or in a nearby device; or somewhere in the cloud. It may be that I've simply read more than my fair share of science fiction, but I don't see any convincing point on the continuum from small artificial pseudo-neuronal structure added to the brain through to having a head full of transmitters and a mind in the cloud where I'd want to deny that it can be part of the mind. An acknowledgement now that external cognitive aids can be parts of the mind, even if only temporary and peripheral, might be good preparation for such a world.

3. Beings without brains could have minds

Imagine that you test your new X-ray specs by peeping inside your two best friends' heads, and discover they have no brains: their heads are full instead of (respectively) electronics and intestines, and they admit that are (respectively) an android and an alien. Since they have given an extensive display of appropriate behaviour over the last decades, you must presumably grant that they have minds (can understand, have some mental states), despite being literally brainless.

This possibility shows at least that the mind is not *identical to* the brain: if there might be cases where there is a mind without a brain, then minds are not identical to brains. Mental states and minds, defined by their functions in cognition rather than their substrate, can be realized in multiple ways. But there is still an interesting sense in which minds could be brains: that of role-filling or office-occupancy. That is, minds may be something like 'whatever substrate centrally serves to enable cognitive processes' or 'whatever is casually most relevant to adaptive behaviour', a role which happens to be filled by the brain in humans (though see above for doubts about whether this is true!). In this case, the mind is the brain for humans, just as the mind is some other organ for your best friends.

My own opinion here is that this 'the mind is the brain, at least for us' way of thinking is not helpful. Consider that when you and your friends want a beer, there are three very different realizations: as a brain process, an electronic process, and so on. As I say of each person 'they want a beer', I don't thereby mean the underlying process, for I have no idea of what those might be, and I mean something very similar with each utterance, while your underlying processes differ wildly. The only way to make out what the three realizations of 'wanting a beer' have in common is not via the processing substrates, which are varied and unknown to me, but through the behaviour that shows a functional description of that mental state is satisfied, and which the three of you have in common.

To return again to a digital computing analogy, the same application (say, a web browser) may run on a variety of different architectures (say, my iOS tablet, my Linux desktop, and my Android smartphone). We may be ignorant of the details of the architectures, or even what and where they are. In order to make out how my talk of having made the same settings across all my devices is anything other rather than three separate and utterly different claims about computing architecture, we must have it that I am talking about the behaviour of all three devices satisfying the same functional description, and not talking about the processing substrates. In the case of you three wanting beers, I know how and when to judge that a person wants an object: all else being equal, they will tend to look at it with interest when it is in sight, reach for it, affirm a want if asked, and so on. I may also guess that you all want beers simply because it is hot and if I were you I would want a beer. Without giving it any conscious thought at all, I may defeasibly infer that you want beers and hold some up for your attention, and so long as your behaviour satisfies the folk psychological script for 'wanting', I properly judge that you all want beers. Insofar as having a mind is connected to having mental states, I thereby properly judge that you have minds on the basis of your behaviours.

Note that in some ways this may seem to be a behaviourist definition of mind, because after all, I properly judge that you have a mind on the basis of only your behaviour. However, the distinctive claim of behaviourism is that reference to mental events or states ought to be avoided entirely (see for example Skinner 1974), or at least subsequently cashed out entirely as behavioural concepts, and we are not attempting to do that. Indeed, the folk psychological script for 'wanting' adverts to a range of other mental states: you will reach for the object only if you *believe* that it is the object you want, and so on. Indeed, as a responsible host, when I hold up the beers I infer you want, I will be sensitive to signs that may reveal subtle conflicts in your wanting: do you want the beer but not want to want the beer (you're a recovering addict), do you not want the beer but want to please me by accepting it (I'm pressuring you to drink), do you want the beer but not want to impose or be obliged to me, would you prefer some other drink, and so on. For all that my ascribing mental states to you is based on your behaviour, the folk psychological script commits me to a huge complex of possible mental states and a network of interactions between mental states that would be anathema to a behaviourist.

If this analysis is correct, to have a mind is to follow or be apt to follow some of these mental state scripts, whether you have a brain or not. Simple minds, for example those of cats, may be apt to follow only a restricted subset of mental state scripts (believing, wanting, fearing, and so on), with even those scripts taking only restricted types of content: my cat can fear a dog, but not that a dog might come tomorrow; can want an unhealthy snack, but not want not to want an unhealthy snack. Fuller minds are apt to follow more of the folk psychological scripts, and are thereby apt for attributions of self-reflection (you regret killing my cat), mental states with abstract conceptual content (you wonder how the body could best be disposed of), reciprocal use of folk psychological scripts (you believe that I am upset), and so on.

In either case, minds are not brains, or even really things at all; they are something like the state of satisfying a certain type of functional explanation and prediction. There is no spooky stuff here. Being say, a racecar, is similar. Any racecar should fulfil a range of functional scripts (can go fast, turn at speed, and so on). Even if the technology of the time means these are invariably subserved or enabled by a gasoline engine, a racecar is not just a gasoline engine. A gasoline engine without the surrounding car is no racecar, and a racecar with an electric motor would still be a racecar. We identify racecars through their speed (and so on) without having to check their engine. Just because we have no noun for the state of satisfying racecar functional script descriptions (having 'racecarness'?) and do have a noun for the state of satisfying mental functional script descriptions (having 'a mind') doesn't mean that there is a thing to be made out in the latter case.

4. Minds and brains have different explanatory levels

The discussion above of how mental state talk is not the same as brain talk, just as computer applications talk is not the same as processing substrate talk, points to a version of the argument against mind-brain identity that does not depend on any intuition that androids and aliens can have minds.

Application talk brings particular objects and interactions to our attention: for example, as a designer, that having only one primary action per screen improves user comprehension, adding comprehensive comments to code makes maintenance easier later, letting users know that you've automatically detected their credit card type reduces support staff load, and so on. Applications running on my smartphone do not have some spooky substance that is distinct from the smartphone, and I'm not a dualist about applications. The application running may be a physical process that centres on the movement of electrons through conducting and semi-conducting materials within my smartphone. But for us to *understand* what makes a good application, *explain* what an application does or how it can be improved, or *predict* what will happen next as we interact with a new application, we need application talk, not substrate talk. The explanatory and predictive fact that there is something on the screen that looks like a button to a user is incomprehensible to a bottom-up view. That is, a complete description of the physical state of the smartphone presumably determines that there is a button on screen, but the massive disjunction of possible states that represent clear buttons is comprehensible as something unitary (as all being about something visibly buttony on screen) only by talking about how the screen will appear to a user and how button-like it is. Users and designers need to understand buttons, and how buttons look to users (and user goals and competing apps and so on), not semiconducting substrates.

Of course, there is nothing unique about application talk. Economics draws our attention to supplies, demands, exchange rates, currencies, and so on; folk psychology draws our attention to beliefs, wants, regrets, and so on. The entities and interactions which are explanatory and predictive depend on the domain. Understanding and predicting currency fluctuations requires talk of market supply and confidence and the actions of banks, not of atoms and electromagnetic fields. Understanding and predicting the behaviour of people requires talk of beliefs and wants, not of dendrites and electrons. Minds are not brains (applications are not patterns of smartphone charge, the US dollar is not a bunch of atoms and EM fields) because they are objects at different explanatory levels.

The reductionist tendency mentioned at the start of this article argues that physical chemistry can be understood in terms of atomic physics, cell biology in terms of how biological molecules work, and organisms in terms of how their component cell systems interact. Minsky (2007) argues that while we cannot yet see how to reduce minds to neuroscience (let alone fundamental physics, presumably), nothing is irreducible: we're just not smart enough to reduce it yet. For Minsky, with enough intermediary explanatory levels between brain cells and high-level mental states (for example, how brain cells combine to make logic gates, logic gates combine to make reasoning modules, modules combine to make mind – perhaps with thousands of levels and modules), we should be able to understand how mind arises from brain.

My own opinion here is that Minsky's view is suitably aspirational, but may ignore in-principle limitations on bottom-up understanding. For example, consider the three-body problem in physics. This is a dynamical system which is chaotic, meaning there is no general analytic solution. Predicting how a particular system will develop essentially requires simulation of the system's development over time, which is computationally demanding, and can yield results which are only ever *arbitrarily* – as against *perfectly* – accurate. If laborious simulation using such numerical methods does not constitute the kind of 'understanding' required for useful explanation and prediction, then the motion of three point masses in space has eluded our understanding despite concerted effort since Newton's 1687 *Principia*.

Neural networks (Minsky's modules, say) can be similarly chaotic. If multiple such modules are linked, as they are in Minsky's model of the mind, the chaotic output of one module (which we can laboriously simulate only to non-perfect accuracy) can form the input of another module, which we can then simulate less accurately as we have non-perfect understanding of its inputs. With even two modules, let alone thousands to comprise just one mind, simulation rapidly becomes computationally intractable. As noted above, we are now struggling to see how the behaviour of *C. elegans* emerges from interactions amongst its 1000 cells. A mouse has about 12 million times as many cells as that, and humans likely have over 37 billion times as many. If there are in principle limits to our understanding of three masses governed by gravity, we might expect in principle limits to our understanding of the development of systems of 37 trillion cells governed by complex electrochemical interactions, the development of which depends on interactions with multiple other systems of 37 trillion cells.

With computationally intractable systems, the quickest way to see how the system will develop is to watch it develop, and the best way to gain an explanatory and predictive handle on it is generally to search for statistical regularities in its behaviour as a system. This is something like traditional weather forecasting: while there is no bottom-up understanding to be had, we may note after long experience that, for example, low pressure zones in the east tend to be followed by rain in the foothills two days later. The laws of physics govern the dynamical system of the weather, without giving us access to emergent regularities of that system. Ultimately, even if Minsky's bottom-up mind approach ends up able to slog out a reductive account of how we can cash out some mental states in terms of neurophysiology, this would be something like being told that a supercomputer has successfully shown how a hurricane developed using molecular motion data. Exciting, and doubtless calling to our attention on the way various fascinating levels of explanation and structures, and a triumph for physicalism over the ancient dualism of Descartes and Leibniz … but not apt to make me give up my talk of pressure zones in the east for molecular motion talk. Understanding, explanation, and prediction need to occur in real time with limited data and computational resources, and a proof that in principle a massively powerful computer will eventually be able to derive my feeling of ennui from a massively accurate scan of me does not help me figure out in real time if you three all want beers.

Fortunately, folk psychology is remarkably successful in explaining and predicting behaviour. Not only can I judge if you all want beers without conscious effort, I regularly make successful predictions that would be simply astonishing from a bottom-up view. For example, you promise that you'll be at my birthday party in a year, and I predict confidently and without particular effort that you will be (we're good friends, and you like my parties). Sure enough, a year later, there you are. Remembering that the bottom-up view has trouble figuring out where three point masses in an idealised space will end up after a finite time, there is simply no way that even a precise understanding of all whatever swathe of molecular motion and charges are taken as the initial condition can be used to predict your location in a year, during which time you may have flown around the world and interacted with thousands of other equally complex systems.

Conclusion

The mind is not a spooky special substance. This recognition seems to encourage us towards a view whereby mental states are merely brain states, and mind is merely the collection of brain states.

However, the mind is not simply the brain. Revisiting the four discussions above:

1. There is no principled way to make out that all and only the brain subserves or enables mental states. If minds are had by anything, its people, not brains.
2. Getting any explanatory handle on the mind (for example, the content of mental states) requires reference to rich social facts. In fact, we might want to allow that parts of the world outside the body (let alone the brain) literally become parts of the mind.
3. Mental states without brain states seem possible. Indeed, mind talk depends on satisfying certain functional scripts, not on brain states. This suggests that minds are not things at all, but something like the state of satisfying a certain type of functional explanation and prediction.
4. Minds and brains have different explanatory levels. Even if all mental facts are determined by brain facts, mental facts are apt to help us in understanding, explaining, and predicting behaviour in a way that even a supercomputer cannot derive from brain facts.

Brain-centred views underestimate the importance of the embodied and social aspects of the mental. Brain-centred views ignore actual and possible (likely near-future) ways that the mental outruns brains. And expecting mental talk to disappear in favour of neuroscientific or physicochemical talk seriously overestimates the actual and in-principle possible (let alone practical) understanding available on a bottom-up view, while seriously underestimating the explanatory and predictive power of mental talk.

References

Burge, Tyler. 1979. Individualism and the Mental. *Midwest Studies in Philosophy*, 4: 73–121.

Descartes, R. 1641. *Meditations on First Philosophy*, in *The Philosophical Writings of René Descartes*, trans. by J. Cottingham, R. Stoothoff and D. Murdoch. Cambridge: Cambridge University Press, 1984, vol. 2.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Macmillan.

Leibiz, Gottfreid Wilhelm. 1714. *Monadology*, in *Monadology and Other Philosophical Essays*, trans. and edited by Paul Schrecker and Anne Martin Schrecker. New York: Bobbs-Merrill Co., 1965.

Minsky, Marvin. 2007. 'Marvin Minsky - What is the Mind-Body Problem?'. https://www.youtube.com/watch?v=DbVClq1pT9M

Putnam, Hilary. 1973. Meaning and Reference. *The Journal of Philosophy*, 70(19), 699-711. doi:10.2307/2025079

Shapiro, L. 2008. Princess Elizabeth and Descartes: The union of soul and body and the practice of philosophy. *British Journal for the History of Psychology*, 7(3), 503-520.

Siegel, Daniel. 2016. *Mind: A Journey to the Heart of Being Human (Norton Series on Interpersonal Neurobiology)*. Norton.

Skinner, B. F. 1974. *About Behaviorism*. New York: Vintage.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*, trans. by G.E.M. Anscombe. Oxford: Basil Blackwell.