

Minds and Material Composition

James Blackmon
Summer, 2010

1.

Physics tells us that the human brain, like any other molecular entity, is a scattered object. Our current scientific theories tell us that this particular scattered object is the thing that is conscious, that realizes or instantiates conscious thought. Somehow, conscious thought arises in this scattered mereological sum of matter as long as this thing is exhibiting certain neural states sufficient for conscious thought. To be more specific, for each mental property that can be instantiated by some brain, there are physical properties which, if instantiated by some brain, provide sufficient conditions for the instantiation of that mental property by that brain.¹ Using a thought experiment, I will argue that this view leads to practical (although not logical) absurdities.² I will then suggest an alternative to this conception of the mind: The mind is more than its material.

2.

Consider some brief but complete mental event in your life, this one you are having now as you read these words complete with all its sensations, thoughts, and

¹ Because I am interested in mental properties as individuated by their phenomenology, I adopt a narrow content view. This is not a rejection of the view that mental properties may also have broad content.

² The thought experiment follows a long tradition of flatly ignoring what may appear to be hard technological limits on what humans can do in order to get entities other than brains to mimic brains, Ned Block's "China Brain" [1978], Daniel Dennett's "Hamlet and Yorick" [1981], and Arnold Zuboff's "Spread Brain" [1981] being but a few.

emotions. Call this mental event *E*. According to the standard view, *E* corresponds to some pattern of neural activation; call it *N*. According to this common view, any brain having *N* will instantiate *E*. Now consider the following question: Is the mereological sum of these neurons exactly that entity which instantiates the mental event in virtue of having the right intrinsic properties?³ Suppose this is the case.

Consider now a system of neurons qualitatively identical to the system that composes your brain while instantiating *N*, but for one neuron, *k*. The difference is that neuron *k* sits inches outside the rest of this brain, and yet it is linked to its corresponding neighbors with circuitry that maintains the appropriate signal transfers. That is to say that despite *k*'s unusual location with respect to the rest of the brain, it behaves neurologically just as it would had it been in its "proper" spot. No doubt, creating such a system would be a fantastic accomplishment. And there is no good reason to place any high bets on this happening in the near future. But the system just described is physically possible, and a complete theory of mind (just as a complete theory of physics) must admit it. Moreover, as part of such a system, neuron *k* could be even farther away, perhaps miles, as long as the signal transfers were reliable and appropriately timed, and even at this distance, *k* would behave as it would had it been in its "proper" spot.⁴

³ Putting the question this way presupposes that mental states are instantiated and that instantiation is a relation between the mental state and some thing. Perhaps there are reasons to avoid adopting this ontology. If so, we can drop that presupposition by asking a more general question: Is the mereological sum's having *N* a sufficient condition for the occurrence of the mental event? But for convenience, I will often speak in terms of instantiation with the understanding that I am not insisting on this ontology. My claims can be interpreted in the more general sense which appeals only to sufficient conditions.

⁴ Time-lag due to distance would not be a problem. Neural signals are much slower than electrical signals or those sent along fiber-optic cables. So, a long fiber-optic circuit could match pace with a short neural one. Also, the transportation of actual neurotransmitters is not required. Stores of the requisite molecules can be on hand at each end of the circuit so that when neuron *k* expels some neurotransmitters that would have gone to a neighbor, a mechanism at the other end of the circuit replicates this by expelling some from its one store. The command to do so would simply be part of the overall signal.

We can now consider an alternate system much like this one but differing in one very significant respect. For, note that the holding of causal relations between neuron k and its neighbors is not a necessary condition for the proper behavior of k . So, instead of one cleverly constructed circuit between k and its neighbors transmitting the requisite signals, an even more cleverly constructed system may exist, one that controls the firings of k and k 's neighbors adventitiously, ensuring that k exhibits the appropriate neurological activity. This system would consist of a master computer (MC) with control filaments running from MC to k and k 's neighbors. No signals would be sent between k and the neighbors. There would not even be signals sent between them *via* MC ; MC takes no input. Instead, MC would simply be programmed to “puppeteer” the neurons, to send the requisite signals out along its control filaments at just the right times so that the entire collection, k included, would instantiate some portion of N , a mere pattern of activations which are now occurring as the collateral effects of MC 's signal-sending activity. The specific pattern of activation that holds of k and any one of k 's neighbors would be achieved by the *projection* of signals from MC onto each neuron, not by the *exchange* of signals between them. Thus, these neurons trigger each other's behavior no more than do the light bulbs of a theatre marquee or the pixels of your computer screen.

There is nothing in physical principle to stop such an arrangement from being applied to all the neurons of a collection. Consider each neuron now causally isolated from its previous neighbors but firing appropriately due to the signals it receives from a filament running to a reprogrammed MC , an MC programmed to direct the firing of

every neuron in the entire collection in just the right way.⁵ This is yet a collection having the pattern N . Call this thing a *de-integrated brain*. The neurons of this de-integrated brain are all neurons causally isolated from each other, but they nevertheless fire just as they would have fired had no causally isolating de-integration obtained.

Now we must confront a question: Is the de-integrated brain a thing that has E , the experience you had previously while reading the beginning of §2? Put another way, if your brain were currently in the de-integrated state but governed by MC so as to exhibit the same pattern of activation it currently has, would you be having the same phenomenal experiences you are having now?

3.

If so, we face a number of very counter-intuitive results that will be outlined here. Recall that the view under consideration is that certain patterns of activation holding of collections of neurons are sufficient for the occurrence of certain mental events, and the thing that instantiates the mental event is just the mereological sum of these neurons. Now for the counterintuitive results.

First, this de-integrated brain may span from Mars to Venus. In fact, there is no limit in principle to how far the neurons may be from each other. Physical instantiations of mental events can be radically strewn across space, their parts bearing little or no causal relevance to each other. Moreover, these physical systems may have less integrity than any cloud of gas, their elements straying far and wide, flying past each other at irregular speeds. The entire system of neurons can do whatever we might imagine as long

⁵ Note that no robust processing is going on. The manner in which the signals go out may all have been arranged ahead of time. In fact, the system for managing such signals would not have to be very elaborate at all, having the same computational complexity profile as the cylinder of a music box.

as each fires in just the right way at just the right time. For some people, this is absurdity enough; physical “mind-having” things are, for some reason or other, intact things. They do not mix or dissipate. But if this is a bad consequence of the view under consideration, it gets much worse.

For our second counter-intuitive result, suppose MC has one extra neuron on hand, and suppose that this neuron is firing in just the same way that your neuron k is firing. We have a new collection of neurons here: your collection, minus k , plus this additional foreign neuron. The collection that is your de-integrated brain drastically overlaps this other collection. In fact, the two collections share all but one of each other’s neurons. Yet, because the two collection-specific neurons behave the same way, both collections exhibit the activation pattern N . Thus, according to the view under consideration, both of these radically overlapping yet different collections have the same experience you just had while reading §2. Of course, MC could have many foreign neurons doing just what k or other neurons in your collection of neurons do, and thus there could be many other de-integrated brains overlapping yours, all having the experience E .

Third, that extra neuron might have come from another de-integrated brain, a collection also controlled by MC or some other master computer. If so, then realizations of mental states just like yours may overlap more than one de-integrated brain. Again, the overlap may involve more than one neuron. Part of one de-integrated brain might be collaborating with parts of other brains in the instantiation of “someone else’s” mental event whether it is one that corresponds to N or not.⁶

⁶ I do not mean to unjustifiably assume that every collection has its own subject. Thus the scare quotes.

Fourth, we must eventually admit that MC along with its control filaments is merely a conceptual aid, one that assisted us in imagining different collections of neurons. Even without it, many collections exist. So, for instance, unless your neuron k is doing something especially rare, then for some period of time there exist many counterpart neurons in the heads of other people (and other animals) and each of these neurons helps to compose a new collection of neurons firing in just the way that your collection of neurons is firing. We do not need MC or cleverly-constructed systems of circuitry in order to establish the existence of collections of neurons that meet the requisite conditions. The consequence is a bizarre (albeit limited) form of panpsychism: Parts of our brains collaborate in instantiating many minds and many of these minds have overlapping physical instantiations.

Fifth, and finally, we must ask whether only neurons may form these collections. On many views, the replacement of one of your neurons with something else that behaves in the same relevant way does not disqualify the resulting system from instantiating the mental state. Indeed, on some views, behaving in the same relevant way as the neuron is plainly sufficient for partaking in the instantiation of the mental state. If so, then our panpsychism runs further amok to involve collections that include other things that act sufficiently like a neuron. Build a single artificial neuron that fires in the right way and you have created multitudes of overlapping physical things that have mental events.

We have reached the end of this horn. If collections of neurons instantiate mental events merely in virtue of having the right patterns of activations, then some form of panpsychism follows.⁷

⁷ Absurd as these consequences may be, I do not see this as a *reductio ad absurdum* for three reasons. First, we obviously have not encountered any specific contradiction, just a view that seems very strange, even

4.

On the other horn of the dilemma, collections of neurons do not instantiate mental events merely in virtue of having the right patterns of activations. One criticism of the thought experiment as it has been developed so far might be that the appropriate kind of neural activation involves more than the right neurons having the right activations at the right times; the activation must *spread* among neurons. According to this objection, the brain certainly is not like a mere theater marquee; appropriate activation necessarily involves real signal transfer between units. If so, then I will seem to have been unfair to those who had this kind of phenomenon in mind. For de-integrated brains do not after all have *N*, if *N* is to be understood in a way that necessitates appropriate signal transfer between units.

No doubt, the distinction must be drawn. I propose to use the term *spread of activation* in the way that involves necessary signal transfer between units. The term *pattern of activation* shall refer (as I have used it) only to the activation states of the neurons over time and remains neutral regarding the cause. Therefore, de-integrated brains exhibit patterns, but not spreads, of activation among their neurons.⁸ The proposal then is to require activation spread, thereby avoiding all the embarrassing panpsychic results just witnessed. Certainly, my previous thought experiment does not involve

absurd. Second, it is possible that some truths are absurd, and if any of the previously covered consequences did hold, we could never know it through any scientific means. Do these collections have full-fledged experiences? This is just the problem of other minds applied to very unfamiliar things, and I admit that for the strict philosophical skeptic, the problem of other minds holds for these things just as it holds for anything else. Third, absurdity alone (without contradiction) is no license for rejection. For, the remaining options may be just as absurd or even more absurd. If all other options turn out to be significantly more absurd, we will have to embrace these absurdities as the most plausible consequences of our standard view.

⁸ The fact that signals are transferred from MC to the neurons in the relevant cases should not be misconstrued as satisfying the condition that signal transfer must hold between the neurons.

collections exhibiting spreads of activation; we have only mere patterns. But does anything other than an *ad hoc* attempt to avoid these consequences support this proposal? If activation spread is required, then why is it required? And how does it work?

Either some appropriate mereological sum is the thing that instantiates the mental event or not. If it is, then whether it instantiates the mental event depends not only on the pattern of activation it has but on whether its neurons exchange signals in the appropriate way. But whether this condition holds will have very little or no physical effect on any of the neurons. Physical changes in any particular neuron are brought about by the immediate event of some signal's reaching that neuron; they are not brought about by the historical event of that signal's having been sent from a legitimate (i.e. neighboring, incorporated, and sufficiently neural) source. So on this view, whether a neuron plays a role in the instantiation of some mental state depends on features of the world that have little or no qualitative effect on the neuron.

One might object that nevertheless when the signal arrives, the identity of its source has had some appreciable effect on the signal, an effect that is in turn the cause of an appreciable effect on the neuron. That is to say that different sources cannot send sufficiently similar signals nor can they bring about sufficiently similar effects in neurons. Perhaps the argument can be made that the neurons of intact brains and de-integrated brains are, for physical reasons, necessarily different at the thermodynamic level and that these differences hold despite similarities in firing patterns. If so, the relevance of such differences still needs to be established, hopefully in a manner that avoids being *ad hoc*.⁹

⁹ As I understand it, this is a nontrivial project, and the faith we place in it will draw heavily from our knowledge of thermodynamics and signal theory. For whether and to what degree there are bound to be

But if relevant differences in signals and neurons do not exist at any sub-neural level and yet signal source matters to whether the mental event is instantiated, then we have a view of consciousness that frankly looks like nothing but magic. The instantiation of a mental event is determined by factors by which the thing doing the instantiating is not affected. For, the de-integrated brain and the intact brain are neuron-for-neuron qualitatively identical; no neuron behaves any differently as a result of its degree of spatial or causal intimacy with another. On this view, mental states do not supervene simply on the material thing. Non-local features of the environment matter, even when they leave no physical trace on the thing that instantiates the mental state.

This is action at a distance of a bizarre kind. A wife in San Francisco becomes a widow when her husband in New York dies. She loses the property *being a wife* and eventually the property will become relevant to future causal processes. However, it is causally irrelevant now. There is no sense in which some omniscient being can discern in just her physical make-up some Leibnizian mark of widowhood. Or, to use a more apt metaphor, this paper, for all you know, may have been written by someone other than the author named. If this is so, then you bear different relations to different people and thereby have different relational properties. Either you have the relational property of *reading a paper by James Blackmon* or you have some other relational property involving reading a paper by another source. But at the moment that you read this, whichever relational property you do have regarding this matter has no causal effect on

differences between the signals from these two different sources (differences that do not affect activation) is an issue of physics, and on that issue this philosophy must defer. This is not to count the problem as a one of pure physics. After all, whether and to what extent any such differences are relevant to the instantiation of a mental state cannot be solved with equations. But I think it is safe to say that this is unexplored territory. Moreover, it would certainly be a drastic shift in cognitive neuroscience to hold that mental properties are not instantiated at the level of neurons and neural activation but at some deeper thermodynamic level, especially when the relevant deep-level differences do not have emergent physical effects on neural activation.

you or your behavior. Science cannot hope to study such properties merely by studying the features or behavior of the things that have them. This is simply because having them is irrelevant to a thing's physical features or present behavior.

Consider one last extension of the thought experiment. Suppose that each of your neurons is causally isolated and governed by MC as before but the neurons are not moved away from each other in any significant way. On the current view in which spread of activation is key, you do not have *E*. Suppose, however, that MC removes these causal barriers and simultaneously cuts its signals. We are supposed to accept that you instantly become conscious as a result, "waking up" in state *E*. And this very significant event occurs even though not a single neuron has diverged in either its state or behavior.

The predicament posed by these consequences deserves scrutiny, but it also deserves loose speculation. I will close with the latter.

5.

Perhaps we can avoid the absurd consequences by rejecting the thesis that it is just the scattered object that instantiates the mental state. We prefer to do this without stipulating any nonphysical entities. One suggestion is that mental states are instantiated by the sum or composition of both the relevant material and whatever it is that keeps it appropriately connected and in "causal contiguity"—that "medium" or "field", we might say, by which the appropriate activations actually spread. Instantiations of minds, on the

view suggested, are physical (in the sense that light, gravity, and space-time is physical), they just are not entirely material in any classical or intuitive sense.¹⁰

We must heed our best physical theories. Whether it makes sense to speak of some medium, some kind of field (presumably at peace with the special-relativistic denial of a luminiferous aether), is largely a matter of physics. But however it is that we can best characterize the transfer of energy or the causal connection between “brute material bodies”, this final attempt to avoid the strange consequences just surveyed appears to have some promise. If it makes sense to say that the signals themselves, along with the material, are collectively the bearers of mental properties, we avoid all so far witnessed absurdities.

Moreover, if it makes sense to say that the signals themselves—and nothing else—are the bearers of mental properties, we also avoid all so far witnessed absurdities. But functionalism at the neural level encourages just this view.¹¹ The material is irrelevant so long as it can perform its proper function, and this function is simply the adequate transmission of energy signals.¹² Thus, unless we are going to cling to the “neuro-chauvinist” dogma that sufficiently neural material is necessary for conscious thought, a commitment to the idea that physical entities instantiate the mental and a commitment to avoiding the absurdities witnessed above together entail a positive thesis: Conscious thought is, or is instantiated by, pure energy, appropriately channeled.

¹⁰ Nor are they entirely material according to a prevalent, surviving view in contemporary science. After all, some parts of minds, on this view under consideration, fill the “empty” regions between and neurons, and these things are not composed of atoms, nor are they commonly believed to have any mass.

¹¹ The neuron replacement thought experiment, attributed to Zenon Pylyshyn [1980], provides conceptual support for functionalism at the neural level. Investigations of this thought experiment begin with Savitt [1982], Cuda [1985], and Chalmers [1996].

¹² To hold that it is the energy transfer, not the matter which enables it, that is relevant is certainly not to deny our mass-energy conservation laws.

Works Cited

- Block, N. 1978. Troubles with functionalism. In C. W. Savage, ed., *Perception and Cognition: Issues in the Foundation of Psychology*. Minneapolis: University of Minnesota Press.
- Chalmers, D. 1996. *The Conscious Mind*. New York: Oxford University Press.
- Cuda, T. 1985. Against neural chauvinism. *Philosophical Studies* 48:111-27.
- Dennett, D. 1981. *Brainstorms*. Cambridge, Mass.: MIT Press.
- Pylyshyn, Z. 1980. The “causal power” of machines. *Behavioral and Brain Sciences* 3:442-44.
- Savitt, S. 1982. Searle’s demon and the brain simulator reply. *Behavioral and Brain Sciences* 5:342-43.
- Zuboff, A. 1981. Hofstadter, R. and Dennett D., eds., *The Mind’s I: Fantasies and Reflections on Self and Soul*. New York: Bantam Books.