

On the Simplicity of Conscious Beings

David Barnett

University of Vermont

Descartes and Leibniz argue that conscious beings are simple, that they have no parts.¹ Though I am not convinced, I do think that simplicity plays a central role in our naïve conception of a conscious being. And I think that failure to appreciate this role has led to confusion over the significance of a number of influential examples in contemporary philosophy of mind. My primary aim here is to argue that our naïve conception of a conscious being does in fact demand simplicity. I do not conclude that conscious beings really are simple, for I am not convinced that their nature is accurately reflected by our naïve conception. A secondary aim is to show that this demand has exerted an unrecognized influence on contemporary philosophy of mind by motivating our reactions to a number of influential examples.

To begin, consider what it might have been like to be Descartes as he wrote the *Meditations*, or to be Hobbes as he fled the Civil War. Now consider what it might have been like to be this *pair* of philosophers during these events. This request is odd, for surely there is nothing it is like to be a pair of people. Pairs of people seem incapable of experiencing. To be sure, two people might have qualitatively identical experiences, even simultaneously. You and I might simultaneously pinch our arms and feel the same sensation, but the pair we form would not feel a thing. The idea that a pair of people might itself feel joy, pain, or fear—that it might be a subject of experience—seems not just unlikely but absurd.

Why? What is it about pairs of people that prevents us from taking seriously the idea that they might be subjects of experience? Here are five salient hypotheses:

- (1) Pairs of people do not have enough parts.
- (2) Pairs of people do not have parts of the right nature.

- (3) Pairs of people do not have parts capable of standing in the right sorts of relations to each other and their environment.
- (4) Pairs of people do not have an internal structure (they are mere sums).
- (5) Some combination of (1) – (4) is to blame.

A sixth hypothesis is that pairs of people *have* parts. Perhaps our naïve conception of a conscious being is that of a simple. Perhaps we find absurdity in the idea that a pair of people might itself be a subject of experience because, when we consider a pair of objects *as a pair*, we cannot ignore the fact that we are dealing with a *composite*.

There is an obvious objection to this hypothesis: If simplicity were truly a core part of our naïve conception of a conscious being, then, for a composite of any sort, we should find absurd the idea that it might be conscious. But consider human organisms. We find nothing initially absurd in the idea that they might be conscious. Indeed, we are inclined to think that we *are* human organisms, even though we are obviously conscious beings and human organisms are obviously composites. Despite this objection, I believe that the hypothesis is correct. Below I defend it by first criticizing its salient rivals (1) – (5) and then responding to the objection.

We begin with (1). On this hypothesis, the idea that a pair of people might itself be conscious strikes us as absurd because pairs of people do not have enough parts. Clearly this cannot be the whole explanation. For consider the quadruplet composed of Descartes, Hobbes, Fermat, and Princess Elizabeth. This sum has twice the number of parts but seems no better a candidate for conscious experience. Or consider the entire world population. Might this huge sum of people itself be consciousness? To be sure, everyone on earth might, by some mysterious force, simultaneously experience the same sensation. But it is absurd to think that the sum of people on earth might *itself* experience any sensation at all. Increasing the number of members does not, on its own, seem to have any downward effect on the degree of

perceived absurdity in the idea that a given sum of people might be conscious. I conclude, then, that (1) cannot by itself explain our original intuition.

What about (2)? On this hypothesis, the idea that a pair of people might itself be conscious seems absurd because pairs of people do not have parts of the right nature. Clearly this cannot be the whole explanation either. For we do not seem to care whether the pair we are considering is a pair of people, a pair of dogs, or a pair of inanimate objects, say, carrots or even neurons. In every case we find absurdity in the idea that the given pair might itself be conscious. To emphasize that we find it absurd and not merely unlikely that pairs of objects might themselves be subjects of experience, consider single members of the preceding pairs: Might a *single* carrot experience pain when someone takes a bite out of it? Though it seems extremely unlikely on empirical grounds, there is at least no initial appearance of a conceptual difficulty in the idea. Might a given *pair* of carrots itself experience pain when someone bites into one or both of its members? Here we sense a conceptual difficulty; the very idea seems absurd. Or consider a single neuron. Might it feel, say, nauseous? Again, it seems highly unlikely, but there is no apparent absurdity in the idea itself. Might a given *pair* of neurons itself feel nauseous? This seems not just unlikely but absurd. Merely varying the nature of the members does not, then, seem to have any downward effect on the degree of perceived absurdity in the idea that a given pair of things might be conscious. I conclude, then, that (2) cannot alone explain our original intuition.

What about (3)? On this hypothesis, the idea that a pair of people might itself be conscious seems absurd because pairs of people do not have parts capable of standing in the right sorts of relations to each other and their environment. What sorts of relations might be relevant? The only remotely plausible candidates are causal-dispositional relations of the sort borne by the parts of an ordinary human brain to one another and their environment. But consider the following scenario. Allowing for some radical changes to the laws of nature, suppose that we travel back in time and shrink Descartes and Hobbes down to the size of Fermat's left and right brain hemispheres, respectively. Suppose that we train the two

philosophers to behave as their respective Fermi-spheres behave and then carefully replace the two hemispheres with the corresponding philosophers. The anesthesia wears off, and we pinch Fermat's right arm (or his *former* right arm, should Fermat not survive the ordeal). When the signal arrives at the top of the spinal cord, Descartes identifies it, notifies Hobbes, stimulates certain surrounding neurons, and moves into a new functional state. As a result, Fermat's head turns and faces his right arm; an irritated look appears on his face; and out of his mouth comes the words, "Stop that!" On a relevant functional level, Descartes does just what Fermat's left hemisphere would have done. And Hobbes does just what Fermat's right hemisphere would have done. Now, given that the causal-dispositional relations borne by Descartes and Hobbes are those that Fermat's two brain hemispheres would have borne, is it any less absurd to think that the pair they form might itself be conscious? To my mind, at least, it is not. How after all could a *pair* of anything itself enjoy conscious experience? To be sure, there is nothing absurd in the idea that *Fermat* might somehow survive the procedure—perhaps he would remain conscious throughout the ordeal. What seems absurd, rather, is that *the pair formed by Descartes and Hobbes* might be conscious. Variation in the relations capable of being borne by Descartes and Hobbes to each other and their environment seems, then, to have no downward effect on the degree of perceived absurdity in the idea that the pair they form might itself be conscious. I conclude that (3) cannot by itself explain our original intuition.

What about (4)? On this hypothesis, the idea that a pair of people might itself be conscious seems absurd because pairs of people do not have an internal structure—they are *mere* sums of their two members. To see that this cannot be the whole explanation, simply consider a structure that is intuitively constituted by, but not identical to, a pair of people. Suppose for instance that, as a polite gesture toward Queen Kristina, Descartes and Hobbes had arranged themselves to form a human throne. Intuitively, the throne would have been constituted by, but not identical to, the pair of philosophers, for the pair, but not the throne, would have survived the subsequent separation of Descartes and Hobbes. Now, were we to be

presented with an ordinary thrown in a non-theoretical context, we would likely reject the idea that the thrown was capable of conscious experience, not because we would detect absurdity in the idea itself, but because we would find the idea highly unlikely given our empirical knowledge of thrown. But consider the thrown constituted by Descartes and Hobbes. Might this thrown be capable of conscious experience? The idea that it might immediately strikes me as not just unlikely but absurd. This, I suspect, is due to the fact that it is impossible to ignore the composite aspect of the thrown. Imposing an internal structure on the entity formed by Descartes and Hobbes seems then to have no downward effect on the degree of perceived absurdity in the idea that what they form might itself be conscious. I conclude that (4) cannot by itself explain our original intuition.

What about (5)? On this hypothesis, some combination of (1) – (4) explains our original intuition. Why do we initially find absurdity in the idea that a pair of people might itself be conscious, but not in the idea that a human organism might be conscious? The answer, on this hypothesis, lies somewhere in the fact that a human organism is a *structured* entity resulting from *billions* of *cells* standing to one another and their environment in a certain complex array of causal-dispositional *relations*, whereas a pair of people is a *non-structured* entity resulting from the *mere existence* of *two* particular *people*. To see that this hypothesis fails, we need to consider the human organism, not as we ordinarily do, as a solid-looking human-shaped animated blob, but *as a structure of billions of cells*—or, better yet, as a structure of trillions of particles. We need to make salient the fact that what we are considering is a *composite*. The more salient we can make this fact, the less comfortable we will be ascribing the possibility of consciousness—until, at its limit, the whole idea will start to seem absurd. My strategy for making this fact salient involves closing the gap between a pair of people and a human organism in stages.

First we eliminate the difference in the *number* of parts. Instead of considering a pair of people, we consider a sum of several billion people. We have seen already that a mere increase in parts has no effect on our original intuition.

Next we eliminate the difference in possible *relations* among the parts. Here we can borrow from any of several influential examples in the literature.

Consider for instance Ned Block's homunculi-head example (1978). We are asked to imagine that a human's head is filled with little men who act in concert to realize the functional states of an ordinary brain. Block never specifies the number of little men. Let us suppose that it is the same as the number of neurons in a typical brain. And let us suppose that the little men bear the same sorts of causal-dispositional relations to one another and to their environment as the neurons of a typical human brain. The idea that this sum of men might *itself* be conscious seems absurd.

Or consider Block's nation of China example (1978). Here we are to imagine that, at a relevant functional level, the members of the population of China come to bear the same sorts of causal-dispositional relations to one another and to their environment as the neurons of a typical human brain. Despite its remarkable new aptitude to function in certain respects like human brain, the idea that the nation of China might itself be conscious seems absurd.

Or consider Putnam's swarm of bees example (1967). Here we are to assume that bees are conscious beings, capable of feeling pain. And we are to imagine that a swarm of them realizes the same functional states as an ordinary human brain—or perhaps as an entire human organism. Elaborating a bit on Putnam's original example, imagine that over the horizon there appears to be a colossal human walking toward us. As it nears, we see that it is in fact a huge swarm of bees. In deciding whether to fire missiles at it, we calculate the projected suffering of each individual bee, but not of the swarm itself, for we do not take seriously the idea that the swarm itself might be capable of experiencing pain.

Block uses his examples to argue against functionalism of the mind. Putnam uses his to argue for a surprising constraint on functionalism. We can use the examples to show that variation both in the relevant sorts of relations borne by the members and in the number of the members has no effect on the perceived absurdity in the idea that a given sum of conscious beings might itself be conscious.

What is the source of the perceived absurdity in these three examples?

Putnam suggests that it lies in the *nature* of the members of the considered sums, more specifically, in the fact that the members are themselves conscious beings. In response to his swarm of bees example, he places a surprising constraint on his functionalistic analysis of pain: “no organism capable of feeling pain possesses a decomposition into parts which separately [are capable of feeling pain]” (1967; p.227). (It is perhaps worth noting how bizarre this move is. Aside from being ad hoc, it runs totally against the spirit of Putnam’s own theory. The very idea behind functionalism is that having a mind does *not* require having parts of any specific nature: so long as something realizes the relevant functional states, it has a mind, regardless of whether it is a simple being, or composed of organic matter, silicon, water, plastic, or *even little tiny men.*)

I doubt that our intuition has to do with the nature of the members of the swarm. The idea of a swarm of conscious bees that is itself conscious seems absurd; but so does the idea of a swarm of *dead* bees, or a swarm of *mechanical* bees. The idea of a *swarm* of anything that is *itself* conscious seems absurd. So long as we consider a given swarm *as a swarm*—thereby never losing sight of the fact that we are dealing with a composite—we will find absurdity in the idea that what we are considering is a conscious being. It makes no difference whether the members of the swarm are squash balls, deviled eggs, planets, people, bees, or even *neurons* (more on this below). Their *nature* matters not; their *number* matters not; and their *relation* matters not.

Unless, that is, we assume from the start that we are composites, as Block does in arguing against Putnam’s hypothesis as to what underlies our relevant intuitions. Block (1978) purports to give a counterexample to Putnam’s hypothesis, an example of a being that is intuitively conscious even though it has parts that are themselves conscious. But his interpretation of the example presupposes that we are composites of a certain sort, namely, human organisms. He asks us to imagine that the sub-atomic particles in our bodies are gradually replaced with functionally equivalent spaceships piloted by tiny

aliens. Block believes, reasonably so, that we might survive such a procedure. But from here he infers that we would literally end up with conscious beings as *parts* at some level, and this inference implicitly assumes that we are *identical* to—and not merely, say, embodied in—our bodies. Ordinarily this assumption would be harmless, but not in the context of searching for the source of our *pre-theoretic* intuitions. Here it only blinds us to certain otherwise plausible hypotheses, for instance, that we resist ascribing the possibility of consciousness to homunculi heads, nations of people, swarms of bees, *and constellations of alien-piloted spaceships* because in each case it is impossible for us to ignore the fact that we are dealing with a composite.

In terms of the nature, number, and relation of parts, then, we have closed the gap between a pair of people and a human organism. In the next and final stage, we explicitly add structure to the object under consideration. To eliminate any question as to whether we have closed the gap completely, we consider the human organism itself.

Extra care is required to keep salient the fact that we are considering a composite. For though we are well aware of the empirical fact that human organisms *are* composites, we are not accustomed to thinking of them *as* composites. In part this is because our visual systems are too coarse-grained to present them to us as composites: we see them, not as structures of trillions of particles separated by vast amounts of space, but as solid blobs. And in part it is because nowhere in the idea of a human organism is the idea of a composite (at least nowhere a priori accessible). By contrast, the idea of a pair of people transparently contains the idea of a composite: a pair of people, by trivial definition, is *composed* of two people. This is why, when we consider a pair of people *as* a pair of people, it is impossible for us to ignore the fact that we are dealing with a composite. In the case of the human organism, however, some effort is required to make its composite aspect salient.

A case invented by Arnold Zuboff (1981) and later adjusted by Peter Unger (1990) will put us on the right track. Unger asks us to imagine that the neurons of his brain are gradually separated from one

another without interrupting communication within his nervous system. The separation proceeds in stages. First his brain is carefully removed from his body and separated into halves: the hemispheres are placed in nutrient-rich vats several miles apart both from each other and from the de-brained body, and radio transceivers are implanted at the interfaces of both hemispheres and the peripheral nervous system. Because radio signals travel at the speed of light, and because ordinary cross-synaptic signals travel at far lower speeds, normal communication within the nervous system can be preserved. In the next stage, the halves are themselves halved: each brain quarter is fitted with transceivers and placed miles from the others. The process is repeated until eventually each neuron sits, miles from the others, in its own container, hooked up to a highly complex radio transceiver. Throughout the procedure the system as a whole maintains its functional integrity. At the final stage it interacts with the body just as it would have had it remained confined to the cranium. The fact that it is now spread out over a large region has no bearing on the evolution of the intrinsic states of its component cells or of those of the de-brained body.

Now let us step back and consider this system of widely scattered neurons. Is it the sort of thing that might itself experience, say, the taste of McDonald's French fries? Might it *be* a subject of experience? I should emphasize that the question is not whether this system might *support* a subject of experience. There seems nothing conceptually problematic in this idea: throughout the procedure *Unger* might remain conscious, and his state of consciousness might depend all the while on the state of the system. What seems conceptually problematic, rather, is the idea that *the system itself might be conscious*. (Note the similarity to Leibniz's windmill example.²)

Unger shares this intuition. He acknowledges, moreover, that the strength of his intuition increases as the scenario plays out. His view as to the source of his intuition is that as the procedure progresses the neurons contribute progressively less, whereas the transceivers contribute progressively more, to the functioning of the system. Why this should matter Unger never says. A better hypothesis, I think, is that the composite aspect of the system becomes progressively more salient.

To test my hypothesis against Unger's, we can make salient the composite aspect of the brain without envisaging any changes to the brain itself. Instead of manipulating the brain, we manipulate our images of it. We imagine, for instance, that we are fitted with a series of magical goggles. Each pair provides a higher resolution image of Unger's body than the preceding pair. Without any goggles, Unger's body appears to us as a solid human-shaped blob. The first pair enables us to see the billions of individual cells that make up Unger's outer layer of skin. The cells are packed so tightly together that body still appears as a solid blob, though one with an intricate pattern on its surface. The second pair is truly magical: it enables us to see all the atoms that make up Unger's body. Because the atoms are separated by relatively enormous regions of space, Unger's body now looks like a scaled-down galaxy of stars. This effect is exaggerated when we don the final pair: it provides us with ultra-fine-grained vision that allows us to see the sub-atomic particles that make up Unger's body. Our visual experience is now very much like it would be if we were gazing into outer-space on a clear night.

With our most powerful goggles on, we ask ourselves whether the system of widely scattered particles before us might itself be a subject of consciousness. If we set aside our empirical knowledge of the correlations between states of human organisms and states of conscious beings, I think most of us will find the idea no less absurd than the idea that a galaxy of stars might itself be conscious. This tells against Unger's hypothesis and in favor of mine.

It also completes my critique of (1) – (5), the salient rivals to my hypothesis that simplicity is a core part of our naïve conception of a conscious being. In all of the hypothetical scenarios we have considered, a composite entity is presented to our minds *as a composite*, and we are asked whether the entity might itself be a subject of consciousness. It matters not whether the entity has two, two thousand, or two billion parts; it matters not whether its parts are people, bees, stars, neurons, or sub-atomic particles; it matters not whether its parts bear the relations typically borne by stars of a galaxy, neurons of a brain, or sub-atomic particles of an entire human organism; and it matters not whether it is a mere sum

or some sort of structured entity—say, a swarm, a constellation, or a brain. What matters is whether it is presented to our minds *as a composite*. If so, we are disposed to find some degree of absurdity in the idea that it might be a subject of consciousness. This suggests that simplicity is a core part of our naïve conception of a conscious being.

What of the objection that we are strongly inclined to think that we are human organisms, even though we are obviously conscious beings and human organisms are obviously composites? My response should be obvious by now: we are not accustomed to thinking of human organisms *as* composites, especially outside philosophical contexts. As I tried to show above, the more salient the composite aspect of human organism becomes, the less inclined we are to ascribe the possibility of consciousness to it—until, at the limit, the very idea begins to seem absurd.³ Clearly we are intimately related to human organisms, but the idea that we—conscious beings—might be *identical* to these highly structured systems of flesh and blood begins to seem absurd as their composite aspect becomes salient.

In arguing that at the core of our naïve conception is the Cartesian view that simplicity is necessary for conscious experience, I hope to have made some progress toward showing that contemporary philosophy of mind is under the influence of this conception in a way that has yet to be recognized. I took advantage of several influential examples from contemporary philosophy of the mind. For the most part, the examples were originally offered as counterexamples to various materialistic theories. They are examples of things—usually hypothetical—that, by the lights of the theory under attack, are capable of conscious experience, even though, intuitively, they are not. To my knowledge, nobody has recognized that these examples have something in common: the entities under consideration are presented in a way that makes it impossible to ignore the fact that they are composites. If a simplicity constraint on our naïve conception of a conscious being is indeed driving our intuitions about these examples,⁴ then, without anyone's knowledge, considerations of simplicity have been driving our intuitions about central examples in philosophy of mind, thereby exerting an unrecognized influence on

the shape of our theories.

Simplicity plays a central role in our naïve conception of a conscious being. Failure to appreciate this role has led to confusion over the significance of a number of influential examples in contemporary philosophy of mind. Whether conscious beings really are simple I do not know. But in any case it is worth acknowledging the role of simplicity in our pre-theoretic conception, for it has a hand in shaping our theoretical views by driving our untutored responses to specific examples.

REFERENCES

- Bennett, J. "On the Simplicity of the Soul," *Journal of Philosophy* 64 (1967): 648-60.
- Block, N. "Troubles with Functionalism." In *Readings in Philosophy of Psychology 1980*, edited by Ned Block, 268-305. Cambridge: Harvard University Press, 1978.
- Chisholm, R. "On the Simplicity of the Soul." *Philosophical Perspectives* 5 (1991): 167-81.
- Descartes, R. "Meditations on First Philosophy." In *Descartes: Selected Philosophical Writings*, 73-122. Cambridge: Cambridge University Press, 1640.
- Leibniz, G. "Monadology." In *Philosophical Essays*, 213-24. Cambridge, MA: Hackett Publishing, 1714.
- Lowe, E.J. "Identity, Composition, and the Simplicity of the Self." In *Soul, Body, and Survival*, edited by Corcoran. Ithaca: Cornell, 2001.
- Putnam, H. "The Nature of Mental States." In *Readings in Philosophy of Psychology 1980*, edited by N. Block, 223-31. Cambridge: Harvard University Press, 1967.
- Unger, P. *Identity, Consciousness, and Value*. New York: Oxford University Press, 1990.
- Zimmerman, D. "Two Cartesian Arguments for the Simplicity of the Soul." *American Philosophical Quarterly* 28 (1991): 217-26.
- Zuboff, A. "The Story of a Brain." In *The Mind's I*, edited by D. and Dennett Hofstadter, D., 202-11. New York: Basic Books, 1981.

¹ See Descartes (1640) and Leibniz (1714). Recent discussion of the issue includes Bennett (1967), Chisholm (1991), Lowe (2001), and Zimmerman (1991).

² Leibniz (1714; section 17) says:

Moreover, we must confess that the perception, and what depends on it, is inexplicable in terms of mechanical reasons, that is, through shapes and motions. If we imagine that there is a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters into a mill. Assuming that, when inspecting its interior, we will only find parts that push one another, and we will never find anything to explain a perception. And so, we should seek perception in the simple substance and not in the composite or in the machine.

³ There is a further reason why, typically, we do not hesitate to ascribe consciousness to human organisms: for practical purposes, questioning the identifications would be disastrous. Imagine always having to clarify whether you are talking about a given person or the corresponding human organism. How inefficient this would be.

⁴ Other influential examples may admit of the same diagnosis. Consider for instance Searle's Chinese room example. Here it is obvious that the person inside the box does not understand Chinese. The only other candidate is the system comprising the person and the box. But, intuitively, this system—an obvious composite—seems to be an

absurd candidate for a subject of understanding—or, more generally, for a subject of consciousness.