NOÛS 42:2 (2008) 308–335

The Simplicity Intuition and Its Hidden Influence on Philosophy of Mind

David Barnett

University of Colorado at Boulder

Some influential intuitions in contemporary debates over the nature of the mind include:

> *Descartes's Zombie:* Bodies physically identical to ours could lack consciousness.
>
> *Huxley's Explanatory Gap:* There can be no explanation of how states of consciousness arise from interaction among a collection of physical things.
>
> *Putnam's Swarm of Bees:* A swarm of bees could not itself be conscious.
>
> *Block's Miniature Men in the Head:* A collection of tiny men realizing the same functional states as an ordinary brain could not itself be conscious.
>
> *Block's Nation of China:* A collection of ordinary people realizing the same functional states as an ordinary brain could not itself be conscious.
>
> *Searle's Chinese Room:* A system comprising a person who does not understand Chinese and a written set of rules could not itself understand Chinese.[1]
>
> *Unger's Zuboffian Brain Separation:* A collection of widely scattered neurons could not itself be conscious.

Some morals that have been drawn from these intuitions include: that physicalism is false (Kripke 1980; Bealer 1994; Chalmers 1996), that Russellian monism, emergentist property dualism, or substance dualism is true (Chalmers 2001), that conscious beings cannot be composed of other conscious beings (Putnam 1967), that machine-state functionalism fails to provide a sufficient condition for consciousness (Putnam 1967; Block 1978),

that instantiating a computer program is not sufficient for understanding (Searle 1980), and that it can be vague whether something is conscious (Unger 1990).

I begin by saying a bit more about the intuitions and the morals drawn (§1). Then I elicit a novel intuition and argue that what best explains it is:

> *Simplicity:* Our naïve conception of a conscious being demands that conscious beings be simple (i.e., that they not be composed of other things) (§2).

I generalize, by arguing that *Simplicity* best explains all of the preceding intuitions (§3). In light of this conclusion about the source of the intuitions, I argue that, unless we are justified in accepting that conscious beings must in fact be simple, the intuitions lack the significance attributed to them (§4).

It is worth emphasizing that I do not argue that conscious beings must *be* simple, only that our naïve conception demands that they be simple. One might admit that our naïve conception demands simplicity, and that this is some evidence that conscious beings must in fact be simple, but maintain that there is countervailing evidence. Whether there is countervailing evidence is not an issue I pursue in this paper. I raise it only to emphasize that those who accept that there is countervailing evidence must reject the principle that conscious beings must be simple *as well as the evidential force of all intuitions based on it*. Thus, if my arguments are sound, then a strong argument that conscious beings need not be simple is a strong argument for dismissing a range of influential intuitions in philosophy of mind.

1. The Intuitions and the Morals Drawn

Descartes' Zombie. Descartes (1641) asks us to consider "the body of a man as a kind of machine equipped with and made up of bones, nerves, muscles, veins, blood and skin in such a way that, even if there were no mind in it, it would still perform all the same movements as it now does in those cases where movement is not under the control of the will or, consequently, of the mind" (Sixth Meditation). The mindless bodies—or *zombies*—that Descartes envisages behave just like ordinary humans *except in those cases where the behavior of the bodies would have been governed by the mind of an occupier, had the body been occupied.* Today's zombies—envisaged by Kripke (1980), Bealer (1994), Chalmers (1996) and others—behave like ordinary humans *without exception*. They are composed of exactly the same matter, which behaves in exactly the same way, as that of their conscious twins. At least initially, zombies of this strong sort seem metaphysically possible. Thus, we have the pre-theoretic intuition *that bodies physically identical to ours could lack consciousness*. Kripke (1980), Bealer (1994), and Chalmers (1994) draw the moral that physicalism is false.

Huxley's Explanatory Gap. T.H. Huxley (1866) remarks, "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of Djin when Aladdin rubbed his lamp." Since Huxley made this remark, there has been significant progress toward explaining the structure and function of the nervous system. But there has been no progress explaining how the nervous system gives rise to states of consciousness, that is, to *phenomenal* states. And it is hard to see how there could be. For our understanding of the nervous system is limited to our understanding of how its parts interact with one another and their environment.[2] Yet what demands explanation is how *any* interaction among a collection of physical things could give rise to a state of consciousness. Understanding the specific mode of interaction realized by the parts of our nervous system is not going to help with this question. Hence the explanatory gap. We have the intuition *that there can be no explanation of how states of consciousness arise from the interaction among a collection of physical things*. Chalmers (1995, 2002) draws the negative moral that physicalism is false and the positive moral that Russellian monism, emergentist property dualism, or substance dualism is true.

Putnam's Swarm of Bees. Hilary Putnam (1967) imagines that a swarm of bees realizes the same (machine) functional states as a human organism. Elaborating on his example, suppose that over the horizon we spot what appears to be a colossal human marching toward us, destroying everything in its path. As it nears, we see that it is in fact an enormous 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 the idea that the swarm itself might experience pain seems absurd. Thus, we have the intuition *that a swarm of bees could not itself be conscious.* Putnam draws the moral that conscious beings cannot be composed of other conscious beings and that (machine state) functionalism therefore fails to provide a sufficient condition for consciousness.

Block's Miniature Men in the Head. Ned Block (1978) asks us to imagine that the head of an otherwise ordinary human is filled with a group of little men. Also inside the head is a bank of lights connected to inbound sensory neurons, a bank of buttons connected to outbound motor neurons, and a bulletin board on which a symbol (designating the current state of the system) is posted. Each man is given a simple set of instructions: if a given symbol is posted, then if certain lights are illuminated, press a given button. In concert, the men realize the same (machine) functional states as an ordinary brain. The idea that this collection of tiny men might itself be conscious seems absurd. Thus, we have the intuition *that a collection of tiny men realizing the same functional states as an ordinary brain could not itself be conscious*. Block draws the moral that (machine state) functionalism fails to provide a sufficient condition for consciousness.

Block's Nation of China. Block (1978) asks us to imagine that the head of an otherwise ordinary human contains only miniature two-way radios hooked up to inbound sensory neurons and outbound motor neurons. The radios send and receive signals to and from citizens of China, who are themselves equipped with two-way radios. In place of a bulletin board, a satellite system displays symbols that can be seen from anywhere in China. Each citizen is given a simple set of instructions: if a given symbol is displayed, then if certain radio signals are received from the sensory neurons, send a given signal to the motor neurons. In concert, the billion or so citizens realize the same functional states as an ordinary brain. The idea that this collection of people might itself be conscious seems absurd. Thus, we have the intuition *that a collection of ordinary people realizing the same functional states as an ordinary brain could not itself be conscious*. Block draws the moral that (machine state) functionalism fails to provide a sufficient condition for consciousness.

Searle's Chinese Room. John Searle (1980) asks us to imagine that someone who speaks only English is locked in a room with two stacks of cards and a set of rules. The cards from the first stack have stories written on them in Chinese. The cards from the second stack have claims about the stories, also written in Chinese. Someone outside the room inserts cards into the room through a slot in the wall; these cards contain questions about the stories, also written in Chinese. The rules, written in English, tell the person in the room how, by the shapes of the Chinese symbols, to identify a claim that appropriately addresses the inserted question. Using these rules, the person inside responds to the outsider by identifying an answer card and sliding it back through the slot. From the perspective of the outsider, it seems that someone inside the room understands Chinese. But the person manipulating the symbols inside the room does not understand Chinese. And it seems absurd to think that the system comprising this person and the written rules might itself understand Chinese, or even be conscious. Thus, we have the intuition *that a system comprising a person who does not understand Chinese and a written set of rules could not itself understand Chinese.* Searle draws the moral that instantiating a computer program is not sufficient for understanding.

Unger's Zuboffian Brain Separation. Peter Unger (1990) asks us to consider a variant of Arnold Zuboff's brain separation example (1981). Unger imagines that the neurons of his brain are gradually separated from one another without interrupting the flow of communication within his nervous system. The separation proceeds in stages. First, his brain is removed from his body and separated into halves: the hemispheres are placed in nutrient-rich vats several miles apart from each other and from the de-brained body; radio transceiver devices 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 several miles from the others. The process is repeated until each neuron sits, miles from any of the others, in a container of its own, hooked up to a 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. Now, is this collection of widely scattered neurons the sort of thing that might itself experience, say, the smell of a rose? Might it *be* a subject of conscious experience? The question is not whether it might *support* a subject of consciousness. Nothing seems problematic about this idea; throughout the procedure Unger might remain conscious, and his state of consciousness might depend all the while on the state of the collection. What seems problematic, rather, is the idea that the collection might itself *be* a conscious being. Thus, we have the intuition *that a collection of widely scattered neurons could not itself be conscious.* Unger takes this intuition to show that the collection of neurons, which he assumes to be conscious before its separation, is not conscious after its separation. He infers that a collection of neurons can be conscious only if it is not scattered too widely. Because what counts as being scattered too widely is a vague matter, Unger draws the moral that there can be vagueness as to whether something is conscious.

2. A New Intuition and Its Source

Ultimately I want to show that *Simplicity* underlies all of the preceding intuitions. But first I want to show that it underlies our intuition about a novel example.

Here is the example. Consider what it might have been like to be Descartes as he wrote the *Meditations*, or to be Hobbes as he fled the English Civil War. Now consider what it might have been like to be this *pair* of philosophers during these events, where this is not intended as an indirect question about the members of the pair, say, what on *average* it was like for the two members. The question is odd; for surely there is nothing it is like to be a pair of people. Pairs of people *themselves* 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. Whereas empirical research would be required to learn more about what it was like to be Descartes as he wrote the *Meditations*, or to be Hobbes as he fled the Civil War, mere reflection on the idea of a pair of people seems to provide a sufficient basis for determining that there was never anything it was like to be the pair comprising Descartes and Hobbes. Thus, we have the intuition *that a pair of people cannot itself be conscious.* Call this *the core intuition*.

Here are four salient hypotheses as to the source of the core intuition:

> *Number:* Our naïve conception of a conscious being demands that conscious beings have more than two immediate parts (and a pair of people has only two immediate parts).
>
> *Nature:* Our naïve conception of a conscious being demands that conscious beings be composed of non-conscious, and *only* non-conscious, beings (and a pair of people is composed of conscious beings).
>
> *Relation:* Our naïve conception of a conscious being demands that conscious beings have immediate parts with the capacity to stand in certain relations to each other and their environment (and a pair of people lacks immediate parts with this capacity).
>
> *Structure:* Our naïve conception of a conscious being demands that conscious beings be structures, that is, things essentially composed of *interrelated* parts (and a pair of people is not a structure: it is a mere *collection* of parts).

Number, Nature, *Relation*, and *Structure* are not rivals. We might have the core intuition because our naïve conception demands that conscious beings be *structures* with a significant *number* of immediate parts of the right *nature* capable of standing in the right sorts of *relations*.

Here is a rival hypothesis:

> *Simplicity:* Our naïve conception of a conscious being demands that conscious beings be simple (and a pair of people is not simple).

One might worry that *Simplicity* is a non-starter. For we have no difficulty entertaining the idea that human bodies—physical systems comprising organs, tissues, cells, molecules, and atoms—are conscious. Indeed, if asked outside a philosophical context whether human bodies—that is, human organisms—are typically conscious, we would, without hesitation, give a positive answer. But if our naïve conception of a conscious being truly demanded simplicity, then, for a composite of any sort, including the human body, we should find absurd the idea that it might be conscious.

This apparent conflict with *Simplicity* is, I believe, only apparent. For we are conditioned by ordinary practice to interpret a claim of the form 'body x is conscious', not in the strong sense, as the claim that x is *identical* to a conscious being, but in the weak sense, as the claim that x is some conscious being's body—or that x is occupied by a conscious being. Ordinarily, when we say that human bodies are conscious, we simply mean that human bodies are occupied by conscious beings. We do not mean that human bodies are identical to conscious beings. For if we meant this, then we would have difficulty entertaining scenarios concerning disembodiment and reincarnation. But we have no such difficulty: Imagine waking up, looking into a mirror, and discovering that you have swapped bodies with someone else, say, Paul

McCartney. No problem. Now imagine waking up and discovering that your favorite desk has swapped bodies with your favorite chair. Big problem: the scenario does not make sense. We cannot make sense of pieces of furniture "swapping bodies" because our conception of them demands that they be *identical* with such bodies. By contrast, we *can* make sense of people swapping bodies because our conception of them does not demand that they be identical with their bodies. Thus, when we say, in a non-philosophical context, that human bodies are conscious, we simply intend to commit to the weak claim that conscious beings *occupy* human bodies.

To emphasize the point, suppose that pieces of furniture *are* conscious, as they are depicted in Disney movies. Then we have no trouble imagining their swapping bodies. But if we have no trouble imagining their swapping bodies on this supposition, then we must be interpreting the supposition in the *weak* sense, as the supposition that conscious beings *occupy* the pieces of furniture. For, interpreted in the strong sense, the supposition is obviously incompatible with body swapping.

Suppose, then, that we explicitly rule out the weak interpretation. We ask: might human bodies be conscious, not in the sense of being occupied by conscious beings, but in the sense of being identical to conscious beings?

Here there may be another merely apparent conflict with *Simplicity*. We might give a positive answer because, as our bodies are ordinarily presented to us, it is easy for us to ignore their composite aspect. On a daily basis, we see our bodies as single, solid, human-shaped blobs. It is obvious that our bodies have left halves, right halves, fingers, hands, arms, and legs; however, because these parts appear to be spatially continuous with one another, the whole body presents itself to our minds, *not* as a system of independently existing parts, but rather as something like an "extended simple." We do not ordinarily see our bodies for what they truly are: structures of organs, tissues, and cells—more fundamentally, structures of quadrillions of tiny particles separated by relatively vast amounts of empty space. As a result, in many ways we are able to think of our bodies as simples. Hume makes a related point:

> An object, whose different co-existent parts are bound together by a close relation, operates upon the imagination after much the same manner as one perfectly simple and indivisible, and requires not a much greater stretch of thought in order to its conception. From this similarity of operation we attribute a simplicity to it, and feign a principle of union as the support of this simplicity, and the center of all the different parts and qualities of the object. (*Treatise* I.iv.6)

By treating our bodies in many respects as simples, we can take seriously the idea that our bodies are *identical* to subjects of experience.

When our bodies are presented to us in a way that makes it impossible to treat them as simples, we resist ascribing consciousness to them. Leibniz illustrates the point:

> 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. (*Monadology*, paragraph 17)

Below we will see that, as we make the composite aspect of a human body more difficult to ignore, our willingness to ascribe consciousness to the human body—and not merely to some being associated with the body—wanes.

Returning to my main line of argument, my case for thinking that *Simplicity* best explains the core intuition begins with a critique of its salient rivals: *Number*, *Nature*, *Relation*, and *Structure*.

Start with *Number*. On this hypothesis, the core intuition is grounded in the principle that conscious beings must have more than two immediate parts. Clearly this cannot be the whole explanation. For consider the quadruplet composed of Descartes, Hobbes, Fermat, and Princess Elizabeth. This collection has twice the number of immediate parts, yet it seems to be no better a candidate for experience. Or consider the entire world population. Might this huge collection of people *itself* be conscious? To be sure, every person on earth might, by some mysterious force, simultaneously experience the same sensation. But it seems absurd to think that their collection might itself enjoy some further experience. To emphasize the point, try to imagine that the humans on earth are all in excruciating pain, while their collection is in a state of pure bliss. This seems absurd. Increasing the number of immediate parts does not, on its own, have any downward effect on the degree of perceived absurdity in the idea that a collection of people might itself be conscious. I conclude that *Number* cannot alone explain the core intuition.

What about *Nature*? On this hypothesis, the core intuition is grounded in the principle that conscious beings must be composed of non-conscious, and only non-conscious, beings. Clearly this cannot be the whole explanation either. For it does not matter whether the pair we consider is a pair of people, a pair of dogs, or a pair of inanimate objects, say, carrots or neurons. In every case, we have the intuition that the pair itself cannot be conscious.

Might a *single* carrot experience pain when someone bites into it? For two reasons, one might answer that, although it seems unlikely on empirical grounds, there is no conceptual difficulty in the idea. First, one might interpret the question as a question about whether a carrot might be *occupied* by a conscious being. Second, one might simply ignore the composite aspect of carrots. But now consider whether a *pair* of carrots might itself experience pain when someone bites into one or both of its members. Here we immediately sense a *conceptual* difficulty; the very idea seems absurd. Neither the weak nor the strong interpretation makes sense: it makes no sense for a single

conscious being to *occupy* a pair of carrots; and it makes no sense for a single conscious being to be *identical* to a pair of carrots. Of course, there might be a conscious being who feels pain whenever one of two carrots is damaged. But the idea that the pair of carrots might itself be such a being seems to involve a conceptual error.

Or consider a single neuron. Might it experience, say, nausea? Again, for two reasons, one might answer that, although it seems unlikely on empirical grounds, there is no conceptual difficulty in the idea. First, one might interpret the question as a question about whether a neuron might be *occupied* by a conscious being. Second, one might simply ignore the composite aspect of neurons. But now consider whether a *pair* of neurons might itself experience nausea. Regardless of what we might discover about neurons—say, that they are actually sophisticated aliens—we seem to know a priori that no pair of them is itself capable of feeling nauseous. Neither the weak nor the strong interpretation makes sense: it makes no sense for a single conscious being to *occupy* a pair of neurons; and it makes no sense for a single conscious being to be *identical* to a pair of neurons. Merely varying the nature of the members of the pair does not, then, have any downward effect on the degree of perceived absurdity in the idea that a given pair of things might itself be conscious. I conclude that *Nature* cannot alone explain the core intuition.

Next consider *Relation*. On this hypothesis, we have the core intuition because our naïve conception of a conscious being demands that conscious beings have immediate parts with the capacity to stand in certain relations to each other and their environment, and a pair of people lacks immediate parts with this capacity. Clearly such relations as *being the brother of* or *standing next to* are useless here. 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. These are the relations that things must stand in if they are to jointly function, on a relevant level, like an ordinary human brain.

But there is no metaphysical obstacle to the possibility of two people standing in any such relation. For illustration, 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. We train the two philosophers to behave as their respective Fermi-spheres and then replace the Fermi-spheres with the corresponding philosophers. 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 outbound 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. At a relevant level, the causal-dispositional relations borne by Descartes and Hobbes are those that Fermat's two brain hemispheres would have borne. Together, Descartes and Hobbes function like an ordinary human brain. Given their new relations to each other and their environment, is it any less absurd to think that the pair they form might itself be conscious? To my mind, it is not. How after all could a *pair* of anything itself be a subject of 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. Thus, variation in how two people are related to each other and their environment has no effect on the degree of perceived absurdity in the idea that the pair they form might itself be conscious. As long as we consider a pair of people *as* a pair of people, we will resist the idea that it is the sort of thing that might be conscious. I conclude that *Relation* cannot by itself explain the core intuition.

What about *Structure*? On this hypothesis, we have the core intuition because our naïve conception of a conscious being demands that conscious beings be structures, and a pair of people is not a structure—it is a mere collection. Whereas a collection of things exists whenever those things exist, a structure of things exists only if those things stand in relations required to *exhibit* the structure. For an example of a structure, consider this clay bowl; for an example of a collection, consider the atoms that constitute the bowl. Intuitively, if we were to spread the atoms evenly about the universe, their collection would survive, but the bowl they now form would not. To see that *Structure* cannot be the whole explanation, 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 Christina, Descartes and Hobbes had arranged themselves in the form of 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. Might this throne be capable of experience? Here we are asked to consider a throne *as a throne constituted by a pair of people*, and the idea that it might be capable of experience strikes us as absurd. Or, consider the brain-like system constituted by the pair of philosophers after they have taken over the function of Fermat's brain. As long as we keep salient the fact that a pair of people constitutes the system, it is hard to see—in the weak or strong sense—how the system itself could be conscious. Imposing a structure on the entity formed by Descartes and Hobbes seems to have no downward effect on the degree of perceived absurdity in the idea that what they jointly constitute might itself be conscious. I conclude that *Structure* cannot by itself explain the core intuition.

Perhaps the core intuition is explained, not by any one of *Number*, *Nature*, *Relation*, or *Structure* alone, but by some combination of the four. For instance, perhaps we have the intuition because the idea of a pair of people

is the idea of a *collection* resulting from the *mere existence* of *two* particular *people*, whereas our naïve idea of a conscious being is that of a *structure* resulting from *many organs*, or *billions* of *cells*, or *quadrillions* of *particles*, standing to one another and their environment in a certain complex array of causal-dispositional *relations*.

One way to see that combining *Number*, *Nature*, *Relation*, and *Structure* does not help is to consider the human body, not as we ordinarily do, as a solid human-shaped animated blob, but as a structure of many organs, or of billions of cells, or of quadrillions of particles. We need to make salient the composite aspect of the body. The more salient we make this aspect, the less comfortable we will be ascribing the possibility of consciousness to the body, until, at the limit, the whole idea will seem absurd. My strategy for making the composite aspect salient is to close the gap between a pair of people and a human body in stages.

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

Next we eliminate relevant differences in *relations* among parts. Here we can employ either of Block's two scenarios: the Miniature Men in the Head, or the Nation of China. In both cases, by virtue of their relations to one another and their environment, a large number of people function, on a relevant level, like an ordinary human brain. And in both cases we have the intuition that the collection itself could not be conscious. Hence, combining *Number* and *Relation* will not suffice to explain the core intuition.

Next we shift our attention from collections to the *structures* they sometimes exhibit. Consider again Block's example of the miniature men in the head. Imagine that the miniature men got inside the head as follows. Very gradually, over a long period of time, every neuron of a healthy human brain was replaced with a miniature, functionally equivalent, man. At the end of the process, billions of miniature men came to constitute a brain-like structure inside the head. Now, there is no problem imagining that the person whose brain undergoes this process survives; the person might remain conscious throughout the ordeal. What seems hard to imagine, rather, is that the structure constituted of the billions of little men might itself be conscious. To my mind, the idea that there might be something it is like to be this structure seems no less absurd than the idea that there might be something it is like to be the collection that constitutes the structure. Shifting our attention from the collection of little men to the structure they exhibit does not seem to make any difference. Combining *Number*, *Relation*, and *Structure* will not, then, suffice to explain the core intuition.

To completely close the gap between a pair of people and a human body, we need to make an adjustment to the nature of the parts of the structure we are considering.

As an initial step in this direction, we can employ Unger's Zuboffian Brain Separation. At the end of the envisaged process, each of Unger's neurons sits, miles from the others, in its own container, hooked up to a complex radio transceiver. The system has maintained its functional integrity; the fact that it is now spread out has no bearing on the evolution of the intrinsic states of its component cells or of those of the de-brained body. Here we have a system comprising a de-brained body and billions of neurons that jointly function, on a relevant level, like an ordinary human brain. This system is similar to an ordinary human body in that it is a *structure*, of roughly the same *number* of parts, of the same *nature*, standing in similar *relations* to one another and their environment as the parts of an ordinary human body. Yet our intuition remains: although Unger might survive the procedure, *the system itself could not be a subject of experience*.

One might worry that due to its partly scattered state the preceding system does not constitute a structure. To address this worry, and to ensure that we have completely closed the gap between our opening pair of people and the human body, we now consider the human body itself.

We can make salient the composite aspect of the body without envisaging any changes to the body. 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, handsome 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 to gaze into outer space on a clear night.

With our most powerful goggles on, we ask ourselves: might the system of widely scattered particles before us itself be a subject of consciousness? Here it is easy to head Hume's warning not to "attribute a simplicity" to this system, and not to "feign a principle of union as the support of this simplicity." It is easy to take the system for what it is: a *structure* of quadrillions of particles. The structure is not some simple object that pops into existence once the particles are so related; at any moment, it *consists* in the particles' being so related. To be sure, there may be a simple object that pops into existence whenever particles are so related, and such an object may be a subject of experience. But, to my mind at least, the idea that this system of particles—considered *as* a system of widely separated

objects—might itself be a subject of experience seems no less absurd than the idea that a galaxy of stars might itself be conscious. I conclude that no combination of *Number, Nature, Relation,* and *Structure* can explain the core intuition.

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 does not matter whether the entity has two, two thousand, or two trillion parts; it does not matter whether its parts are people, carrots, stars, neurons, or sub-atomic particles; it does not matter 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 body; and it does not matter whether it is a mere collection or a structured entity. What matters is whether the entity is presented to our minds *as a composite*. If so, we are disposed to find some absurdity in the idea that it might be a subject of consciousness. Because the idea of a *pair* of people contains the idea of a *composite*, we find absurdity in the idea that a pair of people might itself be conscious. I conclude that *Simplicity* best explains the core intuition.

Before I generalize this conclusion to our opening intuitions, I want to consider two further candidate explanations of the core intuition.

First, we might have the core intuition because our naïve conception of a conscious being demands that conscious beings are *fundamental beings* (and a pair of people is not a fundamental being, for it owes its existence to two other beings). Granted that composites cannot be fundamental beings,[3] this hypothesis does not rival *Simplicity*. For it may be that, by making the composite aspect of an object salient, we make salient the fact that the object is not a fundamental being. While I am sympathetic to the idea that our simplicity intuitions might be explained in terms of a more basic fundamental-being intuition, I cannot explore it here.

Second, we might have the core intuition because, for an object of *any* sort—a collection, a structure, a simple physical thing, a simple non-physical thing, and so on—it is impossible to imagine "from the outside" that object's being conscious. From the outside, we can imagine signs of consciousness, but not consciousness itself. On this hypothesis, we have the intuition that a pair of people cannot itself be conscious because, from the outside, we are incapable of imagining a pair's being conscious. The fact that we lack this capacity has nothing to do with the composite aspect of the pair. For we also lack it with respect to simple things, like electrons and angels. I reject this hypothesis. For at most it could explain why we have the intuition that a given object *might not* be conscious. Yet the intuition to be explained is that a pair of people *cannot* be conscious. The suggested hypothesis is too weak, then, to explain the core intuition.

I maintain that *Simplicity* best explains the core intuition.

3. The Source of the Opening Intuitions

Now I want to generalize, by arguing that *Simplicity* explains all of the opening intuitions.

Our opening thought experiments fit a pattern. First the composite aspect of some object is made salient. Then one of three questions is asked: Might the object exist without consciousness? Could there be an explanation of how consciousness is generated from interaction among the parts of the object? Might the object itself *be* conscious? In each case, we have an intuition consistent with the principle that conscious beings must be simple.

First consider Unger's Zuboffian Brain Separation. We have the intuition *that a collection of widely scattered neurons could not itself be conscious*. What is the source of this intuition? Unger's hypothesis (which he admits being uncertain of) is that, as the procedure progresses, the neurons contribute progressively less, whereas the radio transceivers contribute progressively more, to the functioning of the system. This hypothesis purports to explain why we have the intuition *and* why the intuition gets progressively stronger as the scenario plays out.

A more plausible hypothesis is that as the procedure progresses the composite aspect of the system becomes more difficult to ignore. To test this hypothesis against Unger's, we can make the composite aspect of the human brain progressively more salient without envisaging any changes to the brain itself, thus without relegating any of its functionality to non-neuronal mechanisms. If the intuition remains, that is evidence against Unger's hypothesis and for the simplicity hypothesis. The test has been performed, with our series of Magic Goggles examples, and the intuition remains. I conclude that *Simplicity* provides a more plausible explanation of our intuition that a collection of widely scattered neurons could not itself be conscious.

Next consider Searle's Chinese Room. Here we have the intuition *that a system comprising a person and a written set of rules could not itself understand Chinese*. What is the source of this intuition? Searle suggests that the intuition has its source in a principle concerning the causal powers of brains. He says, "My own view is that *only* a machine could think, and indeed only very special kinds of machines, namely brains and machines that had the same causal powers as brains" (1980; p.519).

Simplicity provides a more plausible explanation. Intuitively, only a subject of experience could be a subject of understanding. And we have seen that no pair of objects—considered *as a pair*—will itself seem capable of experience. Applied to Searle's example, we have the intuition that no pair comprising a person and a written set of rules could itself be conscious; we infer that no such pair could itself understand Chinese. To test this hypothesis against Searle's, we can hypothetically add the relevant causal powers of a brain to a pair of people. If the anti-consciousness and anti-understanding intuitions remain, that is evidence against Searle's hypothesis and for the simplicity

hypothesis. The test has been performed, with our Descartes-and-Hobbes-in-the-Skull-of-Fermat example, and the intuitions remain: intuitively, the enskulled pair of philosophers could not itself be a subject of consciousness or a subject of understanding. If Searle complains that Descartes-and-Hobbes-in-the-Skull-of-Fermat does not have the *relevant* causal powers of a brain, we can replace the parts of an ordinary brain with functionally equivalent conscious beings at a more fundamental level, say, that of the neuron. The test has been performed, with Block's Miniature Men in the Head example, and the intuition remains. If Searle (who is not a functionalist) *still* complains that the envisaged system lacks the relevant causal powers, then we can simply turn to our magic goggles scenario, and consider the brain itself. I conclude that *Simplicity* provides a plausible explanation of our intuition that no system comprising a person and a written set of rules could itself understand Chinese.

Next consider Block's two examples: the Miniature Men in the Head and the Nation of China. We have the intuition *that a collection of people—miniaturized or regular sized—realizing the same functional states as an ordinary brain could not itself be conscious*. Call this the *homunculi intuition*. What is the source of the homunculi intuition?

Block suggests that the intuition is "in part controlled by the not unreasonable view that our mental states depend on our having the [information-processing mechanisms] and/or neurophysiology we have. So something that differs markedly from us in both regards [. . .] should not be assumed to have mentality just on the ground that it is [machine functionally] equivalent to us" (p. 280).

There are at least two problems with this suggestion.

First, the intuition to be explained is that neither system of people *could* be conscious, whereas the view Block attributes as the source of the intuition merely entails that neither system *need* be conscious. If Block were right, we should expect it to seem possible both that the systems are conscious and that they are not conscious. But it does not seem possible that they are conscious; it clearly seems impossible.

Second, the intuition is not sensitive to the status of the view that Block takes as its source. Block himself describes a hypothetical homunculi system that, by Block's own lights, has the same information-processing mechanisms and neurophysiology of an ordinary human brain. He asks us to imagine that the sub-atomic particles in our bodies are gradually replaced with functionally equivalent spaceships piloted by tiny aliens. Our brains would continue to function, down to the sub-atomic level, just as they ordinarily would. (According to Block, our brains would have the same information-processing mechanisms and neurophysiology of ordinary brains in the sense that "No techniques proper to human psychology or neurophysiology would reveal any difference in [us]" (p. 280).) Presumably, we would continue to have experience just as we ordinarily would. The question arises: might we, or any other subject of experience, be *identical* to the envisaged system of aliens?

Might this swarm of alien-piloted spaceships itself be conscious? If we were to view such a swarm with our most powerful magical goggles, our visual experience would be much as it would be were we to witness an invasion of earth by a giant armada of spaceships. The idea that, in addition to the experiences had by the pilots of the ships, there might be a further experience *had by the system comprising the pilots and ships* seems absurd. The homunculi intuition is not, then, sensitive to whether the considered system of people realizes the information-processing mechanisms and/or neurophysiology of a human brain.

Why does Block fail to see that his own elementary-particle-people example undermines his explanation of the homunculi intuition? To Block, it seems that the elementary-particle-people system *would* be conscious. This is because Block assumes from the start that we are composite objects. He says, "Since we know that *we are brain-headed systems*, and that *we* have qualia, we know that brain-headed systems can have qualia" (p. 281). Now, *supposing* that we are *identical* to brain-headed systems, and supposing that our experience would be unaffected by the envisaged alien invasion—as it probably would be—*then* it seems that the elementary-particle-people system would itself be conscious. But the relevant question is not: supposing that we are composite objects, might such-and-such composite object be conscious? Rather, it is: suppositions and philosophical theories aside, might the elementary-particle-people system itself be a subject of experience? To which the intuitive answer is: *no*. Because the homunculi intuition is not sensitive to whether the considered system realizes the information-processing mechanisms and/or neurophysiology of a human brain, the fact that the two systems initially considered fail to realize these features cannot, as Block suggests, be the source of the intuition.

A more plausible explanation of the homunculi intuition is *Simplicity*. One way to make it difficult to ignore the composite aspect of an object is to visualize its parts spread out in space—either by increasing the amount of space between them or by increasing the resolution of our visual experience of them. Another is to substitute for its parts conscious beings, whose individuality is difficult to ignore. This, I suggest, is what is going on with the homunculi examples: because the components of the considered objects are stipulated to be conscious beings, the composite aspect of the considered objects is difficult to ignore, and for this reason we have the intuition that the considered objects cannot themselves be conscious beings. I conclude that *Simplicity* provides a plausible explanation of the homunculi intuition.

Next consider Putnam's Swarm of Bees. Here we have the intuition *that a swarm of bees could not itself be conscious.* What is the source of this intuition?

Putnam suggests that it lies in the *nature* of the members of the swarm, more specifically, in the fact that the members of the swarm are themselves conscious beings. This suggestion is implausible. The idea of a conscious swarm of conscious bees seems absurd; but so do the ideas of a conscious

swarm of *dead* bees, a conscious swarm of *zombie* bees, and a conscious swarm of *mechanical* bees. What difference could it make whether the bees are themselves conscious? The idea of a *swarm* of anything that is itself conscious just seems absurd. It makes no difference whether the members of the swarm are squash balls, deviled eggs, planets, people, bees (dead or alive), or even neurons. The source of the intuition cannot, then, lie in the nature of the members of the swarm, for variation in their nature has little or no effect on the intuition.

A more plausible suggestion is that the intuition has to do with the *composite* aspect of the swarm. It is hard to think of a more obvious example of a composite than a *swarm* of things; part of the very idea of a swarm is that it be composed of other things. 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. Stipulating that the members of the swarm are conscious beings only makes it harder to ignore the composite aspect of the object under consideration. I conclude that *Simplicity* provides a more plausible explanation of the Swarm of Bees intuition.

Next consider Huxley's Explanatory Gap. We have the intuition *that there can be no explanation of how states of consciousness arise from the interaction among a collection of physical things*. What is the source of this intuition?

Chalmers (1995) suggests that the intuition might have its source in a principle concerning an ontological gap between the dispositional properties of physical things and the phenomenal properties of their collections.

The suggestion that our intuition has something to do with the nature of the properties had, or lacked, by the physical things that compose the relevant collection is implausible. For the Explanatory Gap intuition is not sensitive to the sorts of properties we suppose the members of the considered collection to have. We have the intuition that there can be no explanation of how states of consciousness arise from the interaction among a collection of *physical* things, with all their attendant dispositional features. But we also have the intuition that there can be no explanation of how states of consciousness arise from the interaction among a collection of *non-physical* things, with whatever sorts of qualities and dispositions they may have. Indeed, we have the general intuition that there can be no explanation of how states of consciousness arise from the interaction among a collection of *things*, whatever their features may be. Thus, the Explanatory Gap intuition has nothing essentially to do with the sorts of features had or lacked by the members of the considered collection.

A better explanation is that, due to *Simplicity*, we cannot see how a collection of things could be identical to, or constitute, a conscious being. It seems mysterious, then, how states of consciousness could arise out of the interaction among a collection of things. Huxley draws an analogy to a genie arising from a lamp. The analogy is perfect if *Simplicity* is the source of our intuition.

How in the world does a ship arise out of the interaction of a bunch of planks? It is just mysterious. Or not. It is not mysterious because it is easy to see how a ship can be *constituted* by a bunch of planks. It is easy to see how, by relating to one another in certain ways, the planks come to *compose* a ship.

How in the world does a genie arise out of the interaction of the parts of a lamp? Here there truly is some mystery. For the parts of the lamp do not come to *constitute* a genie. Rather, their interaction with one another, together with the hand of Aladdin, somehow *causes* a genie to appear. And it is mysterious *why* such a causal relation obtains.

How in the world does a conscious being arise out of the interaction among a collection of neurons? Here, too, there seems to be some genuine mystery. For, due to *Simplicity*, it is hard to see how a collection of neurons could itself *be* conscious, or how it could *constitute* something that is conscious. But if a collection of neurons cannot constitute a conscious being, then we are left with the mystery of *why* interaction among the neurons *causes* a conscious being to appear. And we are left with the further mystery of why specific interactions cause specific states of consciousness.

Of course, the best explanation of the systematic correlations between states of certain collections of neurons and states of certain conscious beings might be that certain conscious beings are *identical to* or *constituted by* collections of neurons. But this explanation could not be complete. For suppose that certain conscious beings are constituted by collections of neurons. A mystery remains: *how* could a conscious being be the same thing as, or constituted by, a collection of neurons? Or, to state the mystery in terms of *states* of consciousness: *how* could being in excruciating pain be the same thing as having one's parts interact in a given way? The hypothesis that conscious beings *are* composed of other things is not going to help us to see how a conscious being *could be* composed of other things, and so it cannot be a complete explanation of the systematic correlations between states of collections of neurons and states of conscious beings. This is just another face of the Explanatory Gap.

I conclude that *Simplicity* provides a plausible explanation of the Explanatory Gap intuition.

Last, consider Descartes' Zombie. We have the intuition *that bodies physically identical to ours could lack consciousness*. What is the source of this intuition?

Chalmers (1996) suggests that it, too, has its source in a principle concerning an ontological gap between the dispositional properties of physical things and the phenomenal properties of their collections. The idea is that, because there is no conceptual connection between dispositional and phenomenal properties, and because we typically think of physical properties as dispositional properties of physical things, we have the intuition that a body of matter could have any range of physical properties without having any phenomenal properties.

For reasons akin to the ones I gave above, I doubt that the Zombie intuition has its source in a principle concerning an ontological gap between dispositional and phenomenal properties. Just as the Explanatory Gap intuition is not sensitive to the sorts of properties we suppose the members of the considered collection to have, the Zombie intuition is not sensitive to the sorts of properties we suppose the matter of the considered body to have. We have the intuition that there could be bodies—structurally identical to ours, and formed from things that, in terms of their behavior, are dispositionally identical to the things that form our bodies—that have no phenomenal properties. But we also have the intuition that, for a collection of things which themselves have *any range* of intrinsic properties, as well as *any range* of behavioral-dispositional properties, there could be a body, formed from things with all the same properties, that has no phenomenal properties of its own (it may have parts that have phenomenal properties). It is easy to imagine a body of things that is not itself conscious; there is no need to stipulate that the things that form the body have any particular range of properties. Because the Zombie intuition is not sensitive to the sorts of properties we suppose the matter of the considered body to have (aside from alleged properties which *by stipulation* give rise to phenomenal properties of the whole body—e.g., "protophenomenal" properties), its source cannot be a principle concerning an ontological gap between dispositional and phenomenal properties.

Simplicity provides a more plausible explanation of the Zombie intuition. There are two ways to think of human bodies: as systems of independently existing objects; or—as we ordinarily do—as shaped, solid, blobs. If we think of bodies in the first way, as bodies *of things*, then it is not just easy to imagine a body that is not itself conscious; due to *Simplicity*, it is impossible to imagine a body that *is* itself conscious. A body of things might be *associated* with a conscious being, but not *identical* to one. Pre-theoretically, however, we think of bodies in the second way. And we think of ourselves as simple beings that *occupy* our bodies, a bit like genies in lamps. Because we detect no metaphysically necessary connection between ourselves, or any other conscious being, and our bodies, we have the intuition that our bodies, or duplicates of them, could exist *unoccupied*, that is, without any consciousness. Plausibly, the Zombie intuition is just another manifestation of *Simplicity*.

I conclude that *Simplicity* provides a plausible explanation of all of our opening intuitions.

4. The Real Significance of the Intuitions

Supposing that our opening intuitions are manifestations of a more basic simplicity intuition, what morals are we entitled to draw from them?

This depends on the status of the principle that consciousness demands simplicity—for short, the *simplicity principle*. Many reject the simplicity

principle on empirical grounds. The best explanation, they say, of correlations between our states of consciousness and states of certain composite objects—for instance, our brains—is that we are identical to these objects. On this view, even if our simplicity intuitions are some evidence in favor of the simplicity principle, there is countervailing evidence.

Whether there is countervailing evidence is not an issue I wish to pursue here. I raise the issue only to emphasize that, if one accepts that there is countervailing evidence, one must reject the simplicity principle *as well as the evidential force of all intuitions based on it*. Thus, a strong argument against the simplicity principle is a strong argument for dismissing a range of influential intuitions in philosophy of mind. Dialectically, this is an important result, for typically those who are drawing morals from the opening intuitions already accept that there is a strong argument against the simplicity principle.

Begin with Unger's Zuboffian Brain Separation. We have the intuition *that a collection of widely scattered neurons could not itself be conscious*. What is its significance?

Unger draws the moral that there can be vagueness as to whether something is conscious. He assumes that the system of neurons that constitutes his brain is conscious prior to its separation, and he takes the Zuboffian intuition to show that the system is not conscious after its separation. Unger infers that a system of neurons can be conscious only if it is not scattered too widely. Because what counts as being scattered too widely is a vague matter, Unger concludes that, for many points in the scattering process, there is vagueness as to whether the system is conscious. Thus, he draws the moral that "there is little truth in the doctrine that conscious experience is all-or-none" (1990, p. 206).

Unger acknowledges simplicity intuitions at the start of his discussion (1990, p. 5). But he moves immediately to reject the simplicity principle, evidently on empirical grounds: "The descriptions of our examples should be in at least rough conformity with the main outlines of our world view" (p. 8) and "... given the general truth of our view of the world, there is no [...] soul for you to have, nor anything remotely like that" (p. 25).

Of course, if the Zuboffian intuition is explained by *Simplicity*—by the fact that our naïve conception of a conscious being commits us to the simplicity principle—then, on the supposition that the simplicity principle is false, the intuition has no evidential force, and Unger is not entitled to draw the moral that there can be vagueness as to whether something is conscious.

Nor is Unger entitled to draw this moral on the supposition that the simplicity principle is true. For, given the simplicity principle, the Zuboffian intuition is perfectly consistent with the view that there can be no vagueness as to whether something is conscious. What is gradual about the example is not the absence of consciousness, but rather our inability to ignore an aspect of the system that was there all along: its compositeness.

Thus, if *Simplicity* explains the Zuboffian intuition, then, whether or not Unger has sufficient reason to reject the simplicity principle, he is not entitled to draw the moral that there can be vagueness as to whether something is conscious.[4]

What, then, is the real significance of the Zuboffian intuition? At most, the intuition is some evidence in favor of the simplicity principle. Whether it entitles us to endorse the principle depends on the strength of available evidence against the principle.

Next consider Searle's Chinese Room. We have the intuition *that a system comprising a person who does not understand Chinese and a written set of rules could not itself understand Chinese.* What is its significance?

Searle draws the moral that instantiating a computer program is not sufficient for understanding. Searle also rejects the simplicity principle. On his view, recall, "... *only* a machine could think, and indeed only very special kinds of machines, namely brains and machines that had the same causal powers as brains" (1980; p.519).

But if *Simplicity* explains the Chinese Room intuition, then, on the assumption that the simplicity principle is false, the intuition has no evidential force, in which case Searle is not entitled to draw any moral from the intuition.

What is the real significance of the Chinese Room intuition? The intuition can be taken as evidence in favor of the simplicity principle, as well any principle entailed by it. Granted that understanding requires consciousness, and given that the pair comprising the person in the Chinese room and the set of rules instantiates a computer program, we can take the Chinese Room intuition as evidence, in the first instance, that the simplicity principle is true; in the second instance, that instantiating a computer program is not sufficient for consciousness; and, in the third instance, that instantiating a computer program is not sufficient for understanding. Whether we are then entitled to draw any of the corresponding morals depends on the strength of available evidence against the simplicity principle.

Next, consider Block's Miniature Men in the Head and Nation of China examples. We have the homunculi intuition: *that a collection of people—miniaturized or regular sized—realizing the same functional states as an ordinary brain could not itself be conscious*. What is its significance?

Suppose the intuition is explained by *Simplicity*. Block seems to think that, even on this supposition, the intuition qualifies as evidence against functionalism but not in favor of the simplicity principle. He says,

> *No* physical mechanism seems very intuitively plausible as a seat of qualia, least of all a *brain*. Is a hunk of quivering gray stuff more intuitively appropriate as a seat of qualia than a covey of little men? If not, perhaps there is a prima facie doubt about the qualia of brain-headed systems too?
>
> However, there is a very important difference between brain-headed and homunculi-headed systems. Since we know that *we are brain-headed systems,*

> and that *we* have qualia, we know that brain-headed systems can have qualia. So even though we have no theory of qualia which explains how this is *possible*, we have overwhelming reason to disregard whatever prima facie doubt there is about the qualia of brain-headed systems. [. . .]
>
> [A]lthough there is good reason to disregard any intuition that brain-headed systems lack qualia, there is no reason to disregard our intuition that homunculi-headed simulations lack qualia. (1978, pp. 281-82)

Block admits that the homunculi intuition might be an instance of a general anti-physical-mechanism intuition, which, I suspect, he would admit might be an instance of an even more general simplicity intuition. Yet he concludes that we can count the intuition as evidence against functionalism without counting it as evidence against the thesis that physical mechanisms might be conscious.

Block is not entitled to this conclusion. He claims that we have empirical evidence that human bodies are conscious, but not that homunculi heads would be conscious. For the sake of argument, let us grant this. He then infers that we have reason to discount our anti-physical-mechanism intuition as it applies to human bodies but not as it applies to homunculi heads. We cannot grant this. For if our intuitions are truly anti-physical-mechanism intuitions, then the reason we have the specific intuitions (i) *that no human brain could itself be conscious* and (ii) *that no homunculi head could itself be conscious* is that we have the general intuition (iii) *that no physical mechanism could itself be conscious.* Suppose, as Block suggests, that we have empirical grounds to doubt (i). Then, ipso facto, we have grounds to doubt (iii). We thereby have grounds to discount the evidential force of any intuition *based* on (iii), including (ii). Similarly, if our intuitions are truly simplicity intuitions, then empirical grounds to doubt (i) are grounds to doubt (iv)—*that no composite could itself be conscious*, which are grounds to discount the evidential force of any intuition based on (iv), including (ii).

So, contrary to what Block suggests, if we take the homunculi intuition as evidence that functionalism fails to provide sufficient conditions for being conscious, then we must also take it as evidence that simplicity is necessary for being conscious. And if we reject it as evidence that simplicity is necessary for being conscious, then we must reject it as evidence that functionalism fails to provide sufficient conditions for being conscious.

Next consider Putnam's Swarm of Bees. We have the intuition *that a swarm of bees could not itself be conscious.* What is its significance?

Putnam draws the moral that his own functionalist analysis of pain requires a further, non-functionalist, condition, namely, that "no organism capable of feeling pain possesses a decomposition into parts which separately [are capable of feeling pain]" (1967, p. 227).[5] Putnam also rejects the simplicity principle; on his view, pain is a state of a whole organism (1967, p. 226).

But if *Simplicity* explains the Swarm of Bees intuition, then, on the assumption that the simplicity principle is false, the intuition has no evidential force, in which case Putnam is not entitled to draw any moral from the intuition.

Next consider Huxley's Explanatory Gap. We have the intuition *that there can be no explanation of how states of consciousness arise from the interaction among a collection of physical things*. What is its significance?

Chalmers (1995, 2002) draws the negative moral that physicalism is false. As for a positive moral, Chalmers (2001) proposes three candidates: Russellian monism, an emergentist form of property dualism, and substance dualism. The Russellian monist can say that the Explanatory Gap intuition reflects an ontological gap between extrinsic properties of physical things and phenomenal properties of their collections: interaction among a collection of physical things can give rise, by virtue of constitution relations, to ontologically *derivative* phenomenal properties of the collection—properties that are constituted by inaccessible, intrinsic, "protophenomenal" properties of the members of the collections. The property dualist can say that the Explanatory Gap intuition reflects an ontological gap between physical and phenomenal properties: interaction among a collection of physical things can give rise, by virtue of unexplainable psychophysical laws, to ontologically *novel* phenomenal properties of the collection. And the substance dualist can say that the Explanatory Gap intuition reflects an ontological gap between physical and conscious beings: interaction among a collection of physical things can give rise, by virtue of unexplainable psychophysical laws, to phenomenal states of a wholly distinct non-physical being.

Supposing, however, that *Simplicity* explains the Explanatory Gap intuition, we are not entitled to draw any of the preceding three positive morals, or even to conclude that one of the three must be correct.

First consider Russellian monism. On this proposal, consciousness sometimes arises out of interaction among a collection of physical things *because physical things have intrinsic protophenomenal properties that, in the right combinations, come to constitute phenomenal properties of collections of physical things (or of structures constituted by such collections)*. Positing such intrinsic features of physical things brings us no closer to understanding how interaction among the physical things could give rise to states of consciousness. For, as we have seen, our intuition that no collection of *physical* things could itself be conscious has nothing to do with the nature of the members of the collection: we have the general intuition that no collection of *things* could itself be conscious. Positing protophenomenal properties, which by stipulation allow a collection of things with them to itself be conscious, will not help us to understand *how* a collection of things could itself be conscious. If we cannot see how any collection of things could itself be conscious, then we cannot see how—*by virtue of its members' having special intrinsic features*—a collection of physical things could itself be conscious. Hence, Russellian monism does

not even respect the Explanatory Gap intuition. Clearly, then, we are not entitled to draw it as a moral from the intuition.

Second, consider emergentist property dualism. On this proposal, consciousness sometimes arises out of interaction among physical things *because it is a law of nature that a collection of certain sorts of physical things (or a structure constituted by the collection) will itself have certain sorts of experience whenever its members interact in certain ways (with one another and possibly also with their environment)*. But positing this law brings us no closer to understanding how states of consciousness arise from interaction among a collection of physical things. For the law entails the possibility of a collection of physical things which is itself a subject of experience. And we cannot see how there could be such a collection. If we cannot see how it could be that *P,* then we cannot see how it could be a law of nature that *P*. *Given* a constant conjunction between conscious states of collections and certain interactions among the members of the collections, positing a law of nature might help to explain *why* there is such a conjunction. But it cannot help to explain *how*, in the first place, there could be conscious states of collections. Hence, emergentist property dualism does not respect the Explanatory Gap intuition either. Clearly, then, we are not entitled to draw it as a moral from the intuition.

Last, consider substance dualism. On this proposal, consciousness sometimes arises out of interaction among physical things *because it is a law of nature that certain interactions among physical things give rise to certain phenomenal states of wholly distinct conscious beings*. This proposal respects the Explanatory Gap intuition. For there is no problem seeing *how* there could be a nomological connection between the physical state of a collection of physical things and the phenomenal state of some non-physical being. The mystery, on this proposal, is merely one of *why* there should be such a connection, and this is no special mystery, for it plagues all fundamental laws of nature.

Still, the Explanatory Gap intuition does not entitle us to draw substance dualism as a moral. At most, it entitles us to draw the moral that the simplicity principle is true. And substance dualism is not a consequence of the simplicity principle. Given the simplicity principle, states of consciousness that arise from interaction among a collection of physical things must be states of something distinct from the collection. But they need not be states of something *wholly* distinct. They might for instance be states of a simple member of the collection, say, an electron.[6] But then there is room for a monism (one that is closer to Leibniz's than to Russell's) on which our universe contains only conscious simples, which, by virtue of their role in our physical theory, also qualify as physical simples. To be sure, there may be strong grounds to reject such a monism. The point, however, is not that the Explanatory Gap intuition entitles us to such a radical monism, but only that it counts no more in favor of substance dualism than in favor of such a

monism. And so it does not entitle us to draw substance dualism as a moral.

What is the real significance of the Explanatory Gap intuition? We can take it as evidence, in the first instance, that the simplicity principle is true and, in the second instance, that either substance dualism or the preceding non-Russellian form of monism is correct. Whether we are entitled to draw the corresponding morals depends on the strength of available evidence against the simplicity principle.

Last, consider Descartes' Zombie. We have the intuition *that bodies physically identical to ours could lack consciousness*. What is the significance of this intuition?

Kripke (1980), Bealer (1994), Chalmers (1996), and others draw the moral that physicalism is false. But suppose that *Simplicity* explains the Zombie intuition, as follows. We intuit that conscious beings cannot be composites and thus cannot be identical to, or constituted by, bodies of matter. We infer (fallaciously) that conscious beings must be wholly distinct from bodies of matter. Because we detect no metaphysically necessary connection between ourselves, or any other conscious being, and our bodies, we have the intuition that intrinsic duplicates of our bodies could exist *unoccupied*, that is, without any consciousness. This conclusion motivates the Zombie intuition: in fact, certain bodies of matter "have consciousness," in the sense that they are occupied by conscious beings; but, without any difference in their intrinsic features—and thus without any difference in their physical features—they might have "lacked consciousness," in the sense that they might not have been occupied by conscious beings.

If this explanation is right, then the Zombie intuition is an *ill* manifestation of a more basic simplicity intuition. For, as we saw above, the inference from the intuition *that conscious beings cannot be identical to, or constituted by, bodies of matter* to the conclusion *that conscious beings must be wholly distinct from bodies of matter* is invalid. We make the inference only because we fail to consider the (admittedly unlikely) possibility that conscious beings might be simple parts of bodies of matter—electrons, for instance. Due to this failure, it seems metaphysically possible for there to be bodies of matter that are intrinsically identical to ours, but which lack consciousness. Once we correct for the failure, however, this intuition disappears. If we cannot, just by reflection, exclude the possibility that we are simple parts of our bodies of matter, then we cannot, just by reflection, know that intrinsic duplicates of our bodies could exist without consciousness.

The question remains whether *physical* duplicates of our bodies could exist without consciousness. This is the question to which our initial Zombie intuition afforded a positive answer. In light of our discussion about the source of this intuition, our attitude toward this question will depend on our attitude toward some difficult issues about the nature of physical properties. Suppose for illustration that you are an electron that is part of the body of matter to which, in an ordinary context, we would point to refer to you.

Because (i) you are conscious, (ii) at least some of your states of consciousness are intrinsic, and (iii) you are part of this body of matter, it is impossible for there to be an intrinsic duplicate of this body that lacks consciousness. Might there, nonetheless, be a mere *physical* duplicate that lacks consciousness? This depends on whether being physically identical to a conscious electron entails being conscious. And this, in turn, depends on which features of a conscious electron would qualify as physical.

Take for example charge. Given that all electrons have charge, and that charge is a paradigmatic physical property, any physical duplicate of a conscious electron would have charge. Does having charge entail being conscious? This depends on what charge is.

On one live view, charge is *whatever intrinsic feature of particles is actually playing the "charge role" in our physical theory.* On this view, if the occupant of the charge role is phenomenal, then being charged entails being conscious; if it is not, then being charged might not entail being conscious. So, generalizing over all physical properties of electrons, on one live view of physical properties it is an open question whether, necessarily, a physical duplicate of a conscious electron would be conscious.

On a rival view, charge is *the second-order property of having some intrinsic property or other that plays the "charge role" in our physical theory*. On this view, even if the charge role is actually occupied by a phenomenal property, it might not have been, and so being charged does not entail being conscious. Generalizing over all physical properties of electrons, on a second live view of physical properties a physical duplicate of a conscious electron might not be conscious.

Thus, whether one retains the Zombie intuition in light of recognizing the possibility that conscious beings might be simple parts of bodies of matter will depend on what view one holds of the nature of physical properties.

What, then, is the real significance of the Zombie intuition? Given that it is an ill manifestation of a simplicity intuition, in the first instance it might provide some indirect evidence for the truth of the simplicity principle. Then, depending on the status of the debate over the nature of physical properties, in the second instance it might provide some evidence against physicalism. Whether this evidence entitles us to draw the corresponding morals depends, once more, on the strength of available evidence against the simplicity principle.

5. Conclusion

Some influential intuitions about the mind include: Descartes's Zombie, Huxley's Explanatory Gap, Putnam's Swarm of Bees, Block's Miniature Men in the Head and Nation of China, Searle's Chinese Room, and Unger's Zuboffian Brain Separation. Some morals drawn from these intuitions include: that physicalism is false (Kripke 1980, Bealer 1994, and Chalmers

1996), that Russellian monism, emergentist property dualism, or substance dualism is true (Chalmers 2001), that conscious beings cannot be composed of other conscious beings (Putnam 1967), that machine-state functionalism about consciousness is false (Putnam 1967, Block 1978), and that there can be vagueness as to whether something is conscious (Unger 1990).

I elicited the intuition that a pair of people cannot itself be a subject of experience. I argued that the source of this intuition is a naïve commitment to the principle that conscious beings must be simple. I generalized, by arguing that this is the source of all of the opening intuitions. In light of this result, I argued that, with respect to some of the opening intuitions, we are not entitled to the proposed morals; and, with respect to others, we are only entitled to the morals on the condition that we are justified in accepting that conscious beings must be simple.

If my arguments are sound, then—unless we are willing to take seriously the principle that conscious beings must be simple—we must dismiss a range of influential intuitions in philosophy of mind.[7]

Notes

[1] This is just one of several intuitions elicited by Searle's Chinese Room example. Searle himself focuses most of his attention on a different intuition, namely, that the person inside the room does not understand Chinese despite his ability to manipulate the Chinese symbols in the relevant ways.

[2] To be sure, neuroscience can provide a certain sort of explanation: it can provide compositional mappings of phenomenal states onto neural states; for an example of how these mappings proceed in the case of visual experience, see Bechtel 2001. But the defender of the gap will say that these are not the kind of explanations she is looking for.

[3] Not everyone will grant this. Lynne Rudder Baker, for instance, holds that persons are composites, yet wants "to avoid the reductionism of those who think either that persons are not among the fundamental kinds of things that (now) exist or that everything that exists (including persons) can be fully understand [sic] in subpersonal terms" (2000, p. 22). I am unclear what Baker means by 'fundamental'; to me it seems contradictory to say, of something, both that it is composed of other things and that it is fundamental.

[4] This problem may be an instance of a more general problem with Unger's methodology. Unger begins his book by rejecting the simplicity principle yet admitting that we have simplicity intuitions. He then proceeds to argue against a number of theses that, plausibly, are intuitive in the first place only because they appear to be required by the simplicity principle. For instance: "A subject is in no way a matter of degree" (1990, p. 41.); "A subject is separate and distinct from every other subject" (*ibid*); and "A subject is absolutely indivisible" (*ibid*). For each thesis, Unger constructs a hypothetical scenario to serve as a counterexample. We are asked to interpret the examples on the assumption that the simplicity principle is false. But this assumption renders the examples superfluous if the intuitive appeal of the theses under attack relies, in the first place, on the simplicity principle.

[5] Note how odd this move is. Not only is it ad hoc, it runs against the very spirit of the functionalist proposal that Putnam is in the midst of defending. The 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 composed of organic matter, silicon, water, plastic, or even *conscious beings*.

[6] Cf. Chisholm 1991.

[7] For helpful comments and discussion, I am grateful to Yuval Avnur, Chris Heathwood, Mike Huemer, Mark Moyer, Thomas Nagel, Adam Pautz, Derk Pereboom, Dave Robb, Rob Rupert, Peter Unger, and to an anonymous referee for this journal.

REFERENCES

Baker, Lynne Rudder. (2000) *Persons and Bodies: A Constitution View*, New York: Cambridge University Press.

Bealer, George. (1994) "Mental Properties," *Journal of Philosophy*, 91, pp. 185–208.

Bechtel, William. (2001) "Decomposing and Localizing Vision," In *Philosophy and Neuroscience*, edited by W. Bechtel, P. Mandik and J. Mundale, Malden, Mass.: Blackwell, pp. 225–49.

Block, Ned. (1978) "Troubles with Functionalism," in *Readings in Philosophy of Psychology* 1980, edited by N. Block., Cambridge: Harvard University Press, pp. 268–305.

Chalmers, David. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.

———. (2001) "Consciousness and its Place in Nature," in *Philosophy of Mind: Classical and Contemporary Readings* 2002, edited by D. Chalmers, New. York: Oxford University Press, pp. 247–72.

Chisholm, Roderick. (1991) "On the Simplicity of the Soul," *Philosophical Perspectives*, 5, pp. 167–81.

Huxley, Thomas Henry. (1866) *Lessons in Elementary Physiology*.

Kripke, Saul. (1980) *Naming and Necessity*, Cambridge: Harvard University Press.

Putnam, Hilary. (1967) "The Nature of Mental States," in *Readings in Philosophy of Psychology 1980*, edited by N. Block., Cambridge: Harvard University Press, pp. 223–31.

Searle, John. (1980) "Minds, Brains, and Programs," *The Behavioral and Brain Sciences*, 3, pp. 417–24.

Unger, Peter. (1990) *Identity, Consciousness, and Value*, New York: Oxford University Press.

Zuboff, Arnold. (1981) "The Story of a Brain," in *The Mind's I*, edited by D. Dennett and D. Hofstadter., New York: Basic Books, pp. 202–11.