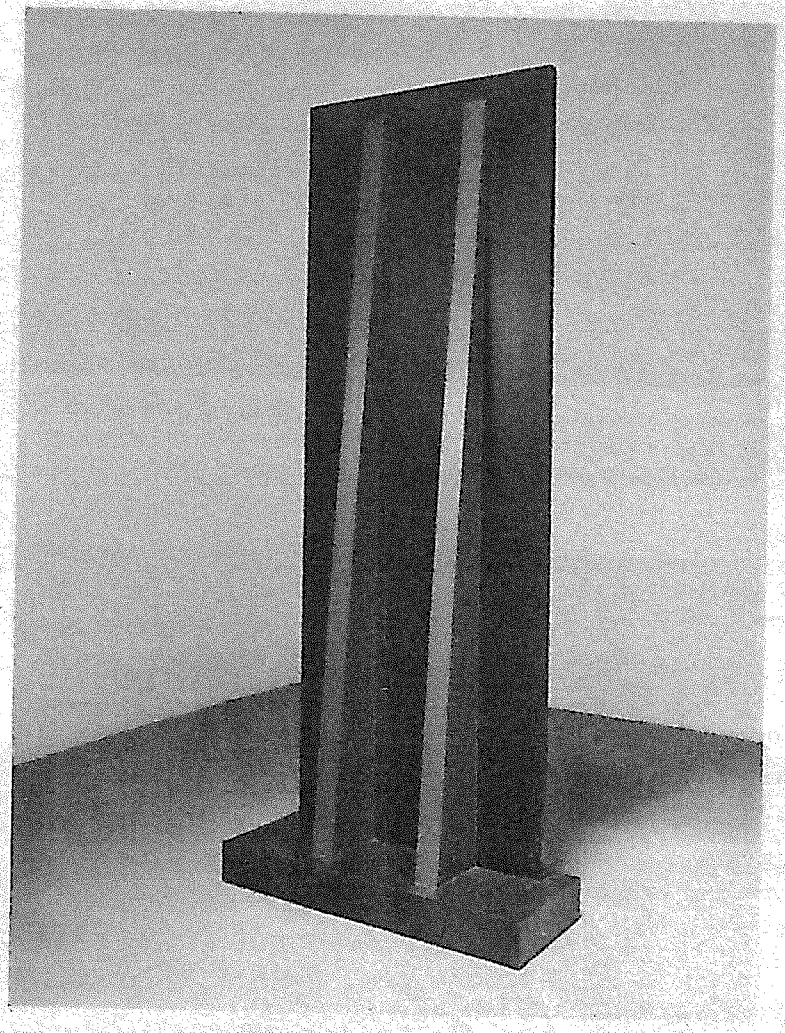

FIGURE 19 Anne Truitt, *Hardcastle*, 1962. Acrylic on wood. 99 ¾ × 42 × 16 inches. Private collection © annetruitt.org / Bridgeman Images.

CHAPTER 12

The Metaphysics of Transhumanism

Eric T. Olson

"Uploading"

Transhumanism is a movement aimed at enhancing our lives by means of futuristic technology. The name derives from the ultimate goal of overcoming the limitations imposed by our humanity. Illness, injury, hunger, exhaustion, ageing, and death set a limit to the length and quality of a human life. There's only so much you can do to make a human being better off, simply because of what it is to be human. But if we could cease to be human—or better, cease to be biological at all—we could free ourselves from these constraints.

Transhumanists hope to achieve this by what they call "uploading." The term is tendentious. *Real* uploading is copying digital data—a document or a video, say—from an ordinary computer to a central web server. I copy my lecture notes from my desktop computer to the university server, for example, enabling my students to make their own copies on their computers at home. Transhumanist uploading involves much more than this. First, the psychological information in your brain is read by a scanning device (erasing it from the brain in the process) and converted into a digital format. It's then transferred to a

Eric T. Olson, *The Metaphysics of Transhumanism* In: *Human*. Edited by: Karolina Hübner, Oxford University Press. © Oxford University Press 2022. DOI: 10.1093/oso/9780190876371.003.0019

computer: that's the part of the process that is properly called uploading. But this information is not merely stored on the computer in the way that my lecture notes are. Rather, it's used to create an electronic person with your personality, knowledge, preferences, plans, memories, and other mental features. This person is not made of flesh and blood but is "realized in" or "implemented on" the computer. The information gathered from your brain is programmed into the computer, and this creates a conscious, thinking being there.

But the aim is not merely to create electronic beings psychologically like ourselves, impressive though that would be. Rather, we ourselves are to move from the flesh to the digital realm. The process is literally supposed to transfer a human being to a computer.

Once that has happened, we shall be entirely inorganic and thus immune to illness, exhaustion, ageing, and death. If the machines that house us are damaged, we can move to other machines by a simple electronic data transfer. Travel will be as easy as emailing. No one will need food, shelter, clothes, or furniture. Our intelligence, memory, patience, capacity for pleasure, and physical strength (if we get robotic bodies) can be enhanced indefinitely—as can the length of our lives. Our human limitations will be a thing of the past. Transhumanists see this as a great benefit. (Whether they're right about that is an important question, but not one I'll try to answer.)

Uploading offers a secular version of traditional beliefs in the afterlife. It promises eternal life without any troubling religious elements: God or an immaterial soul, for instance.[1] Transhumanists have more modest goals as well, but this one is the most interesting.

1 Stephen Hawking once said: "it's theoretically possible to copy the brain onto a computer and so provide a form of life after death" (Nick Collins, "Hawking: 'In the Future Brains Could Be Separated from the Body,'" *Telegraph*, September 20, 2013). Of course Hawking meant, "copy *the psychological information from* the brain." I thank Melanie Challenger for this reference.

All this is founded on the assumption that advanced technology will make it possible to transfer a person from a human body to a computer. As the leading transhumanist Nick Bostrom puts it: "if we could scan the synaptic matrix of a human brain and simulate it on a computer then it would be possible for us to migrate from our biological embodiments to a purely digital substrate (given certain philosophical assumptions about the nature of consciousness and personal identity)."[2] Bostrom has no doubt that we *could* "scan the synaptic matrix of a human brain and simulate it on a computer": not today and not tomorrow, but one day.

You may not share Bostrom's confidence in the power of technology. You may even wonder whether "uploading" is physically possible. But there are deeper worries. He concedes that the story presupposes "certain philosophical assumptions about the nature of consciousness and personal identity." Transhumanists are confident that these assumptions are true. But should they be?

What Uploading Presupposes

Bostrom's "assumption about the nature of consciousness" is that there can be genuine artificial intelligence. It's not only possible for computers to behave intelligently—to play chess and recognize spam emails, for instance—but actually to *be* intelligent and conscious. They can have all the mental properties that you and I have.

Philosophical tradition calls a thing that is capable of rational thought, conscious experience, and awareness of itself as a being distinct from others a *person*. A person is an intelligent, self-conscious being. We human beings are people in this sense (or "persons," as the lawyers

2 Nick Bostrom, "What Is Transhumanism?," last modified 2001, http://www.nickbostrom.com/old/transhumanism.html.

say), and Martians of sufficient intellect would be too. The uploading claim presupposes that it's possible to create a purely electronic person simply by programming a computer in the right way. Computers, in other words, could be people—or at any rate an electronic person could exist in or on a computer. Call this the *AI assumption*.[3]

The AI assumption does not imply that electronic people could be *human*. They certainly wouldn't be biologically human: the whole point of uploading is to liberate us from biology. Whether they would be human in a psychological sense is left open.

So transhumanist uploading presupposes that there could be an electronic person. But it also presupposes that a human person could *become* an electronic person. We could get *you* into a computer, not merely create a new person there with your mental properties and memories of your life—a psychological replica of you. The *same* person could be first "in" a human organism and then "in" a computer. Bostrom calls this an "assumption about personal identity" because it has to do with what it takes for a person to continue existing, as opposed to ceasing to exist and perhaps being replaced by someone else who didn't exist before.

The claim is easy to misunderstand. It's not that someone's "consciousness" could be transferred to a computer. That's a dark statement. Consciousness is a property: it's the one had by conscious beings, just as volume is the property had by things extended in three dimensions. What would it be to move a property of yours—your volume, say, or your temperature—to a computer? It might mean giving the computer that property: changing the computer so that it has the same volume or temperature that you have. But the claim that we could give computers the property of being conscious was the AI assumption. The claim at issue now is not just that there could be a conscious being in the

3 Bostrom calls it an assumption about *consciousness*, as opposed to intelligence or mentality generally. This is because he assumes either that all mentality requires consciousness or that computers could have any mental properties that don't require it. Both points are contested.

computer, but that this conscious being could be you. Describing this as "transferring your consciousness" is an obfuscation.

If we speak of moving a person or a student or a philosopher from a human organism to a computer, we see the difficulty straightaway: how can you upload a thing made of flesh and blood? But this difficulty is obscured if we speak of transferring one's consciousness. There is no obvious obstacle to transferring someone's consciousness to a computer, simply because the state of affairs that description evokes is so vague. It may help break the spell of the word "consciousness" to replace it with its synonym *sentience*. No one would speak of transferring someone's sentience to a computer. Talk of uploading "the mind" or "the self" is equally opaque. The personal-identity assumption is that a human person can move from an organism to a computer by uploading.

This presupposes the AI assumption but doesn't follow from it. If we could become electronic people, then electronic people must be possible. But that possibility would not imply that a human person could become an electronic one. Suppose, by analogy, that it were metaphysically possible for there to be a god—an immaterial, supernatural person. That would be no reason to suppose that we could become gods ourselves.

The Personal-Identity Assumption

Transhumanists have eagerly defended the AI assumption—the possibility of electronic people. But they have said little about the further claim that we could literally move to computers and become electronic people ourselves—the personal-identity assumption. This is the claim I want to examine.[4] Why should we believe it?

The transhumanists' answer is that it follows from an attractive view of personal identity. By this they mean what it takes for a person to

4 I discuss the AI assumption in Olson, "The Metaphysics of Artificial Intelligence," in *Consciousness and the Ontology of Properties*, ed. Mihretu Guta (London: Routledge, 2019).

persist from one time to another—to continue existing rather than cease to exist. What sorts of adventures is it possible, in the broadest sense of the word "possible," for you to survive, and what sort of event would necessarily bring your existence to an end? What determines which past or future being is you? Suppose you point to a child in an old class photograph and say, "That's me." What makes you that one, rather than one of the others? What is it about the way she relates then to you as you are now that makes her you? For that matter, what makes it the case that anyone at all who existed back then is you? This is the question of personal identity over time.[5]

Transhumanists claim that the persistence through time of any person, human or otherwise, consists in a sort of psychological continuity. For a future being to be you is for its mental properties then to depend causally in a certain way on the mental properties you have now—or for there to be a chain of such causal connections. In order to be you, a future being need only inherit your personality, memories, beliefs, preferences, plans, capacity for consciousness, and so on in the right way.[6] This inheritance can take any form, as long as it's reliable. It doesn't require the continuous functioning of your brain.

Think of the "transporter" from the television series *Star Trek*. When the Captain has finished his business on the alien planet, the machine scans him, dispersing his atoms. The information thereby gathered is then sent to the starship *Enterprise*, where it's used to assemble new atoms precisely as the Captain's were when he said "Beam me up!" The result is someone both physically and mentally just like the Captain. And not by accident: the machine reliably produces this result. (Otherwise no one would use it.) The story tells us that this suffices for the man who materializes on board the ship to be the Captain

5 Beware: the phrase "personal identity" is used to mean many other things besides our continued existence. I prefer to call it the *persistence question*.

6 For details, see Sydney Shoemaker, "Personal Identity: A Materialist's Account," in Sidney Shoemaker and Richard Swinburne, *Personal Identity* (Oxford: Blackwell, 1984), 90.

himself—not merely a perfect copy, mistakenly convinced that he's the Captain.

Or consider Sidney Shoemaker's "brain-state transfer" machine (equally fictional).[7] It scans your brain as in the uploading story, recording and then erasing the psychological information stored there. This information is then transferred to another human organism with a "blank" brain, again resulting in someone who is psychologically just as you were at the time of the scan. Shoemaker argues that because this being would be psychologically continuous with you, he or she would be you: the process would move you from one organism to another.

The claim, then, is that our continued existence consists in psychological continuity of this sort. Derek Parfit once called this the "wide psychological criterion" of personal identity.[8] It implies that if the person who appeared in the computer as a result of uploading the information from your brain were psychologically continuous with you (supposing, as the AI assumption says, that there could be such a person), she would be you.

It's hard to see how uploading could move you to a computer unless the wide psychological criterion is true. But whatever its appeal to transhumanists and Trekkies, not many philosophers accept it. Some deny that our persistence consists in psychological continuity of any sort.[9] And most of those who think it does consist in psychological continuity require this continuity to be achieved by the "continuous physical realization" of our mental properties.[10] You could perhaps leave your body behind by having your brain removed and kept intact, but not merely by having its information read off and then copied somewhere else. Think about the period between the brain scan and

7 Shoemaker, "Personal Identity," 108–11.

8 *Reasons and Persons* (Oxford: Oxford University Press: 1984), 207. He does not actually endorse this view.

9 I'm one of them: see Eric Olson, *What Are We?* (New York: Oxford University Press, 2007), 23–44.

10 The phrase is from Peter Unger, *Identity, Consciousness, and Value* (New York: Oxford University Press, 1990), 140.

the creation of a person like you in the computer (or between your "departure" by transporter and the appearance of someone like you at the destination). During that time your mental properties are not "realized" at all.[11] You have no mental properties then: no thoughts, experiences, plans, memories, or preferences. You lack even the capacity for mental activity—and not because that capacity is temporarily disabled, as by a general anesthetic, but because there is no mechanism whose normal workings would enable you to exercise it.

There are grave objections to the wide psychological criterion, and thus to the metaphysical possibility of uploading. Here are just two.

The Mystery of Branching

The first is well known. If you could be uploaded once, you could be uploaded twice. The relevant information could be read off your brain and copied simultaneously to two independent computers in just the way that transhumanists would copy it to one. The result would be two electronic people, each psychologically just as you were when you were scanned. Each would get his or her mental properties from you in the same reliable way. If that suffices for someone to be you, then both must be.

But that's logically impossible. One thing can't be identical to two things that are distinct from each other. If you and the first electronic person were one, and you and the second electronic person were one, then the first electronic person and the second electronic person would also have to be one. This is an elementary fact about the numerical concepts *one* and *two*. Yet there are two people afterward, not one. Supposing that you move to two computers by "double upload" leads to a contradiction.[12]

11 Shoemaker ("Personal Identity," 110–11) says they are realized in the machine. But this is realization in a different sense from the one most psychological-continuity theorists speak of.

12 Temporal-parts theorists say otherwise. This is a large topic that I cannot go into here; for details see David Lewis, "Survival and Identity," in *The Identities of Persons*, ed. Amélie Rorty (Berkeley: University of California Press, 1976); Olson, *What Are We?*, 117–19. I discuss its application to uploading in Olson, "The Central Dogma of Transhumanism," in *Perspectives on the Self*, ed. Boran Bercic (Croatia, Rijeka: University of Rijeka Press, 2017).

This sort of problem is much discussed because it arises on almost any psychological-continuity view of personal identity. The most commonly proposed solution is to say that someone's being psychologically continuous with you in the future is not actually sufficient for you to continue existing. What's required is *nonbranching* psychological continuity. A future person is you only if she is psychologically continuous with you, and in addition this continuity does not take a "branching" form where there are two such people.[13] So transhumanists may say that if the psychological information from your brain were uploaded only once, the resulting person would be you; but if it were simultaneously uploaded twice, neither resulting person would be you. Each would be a new person with false memories of your life—memories of things she never did. You could move to a computer by single upload but not by double upload.

But this creates a mystery. Why should an event that would normally preserve your existence bring it to an end if accompanied by a second such event—one having no causal influence on the first? What is it about the second upload that destroys you?[14] No satisfying answer to this question has ever been proposed. It's no good saying that surviving double uploading would lead to a logical contradiction. That may be a reason to think that it *is* impossible, but no explanation of why it is. It doesn't explain how the second upload brings your existence to an end.

The current proposal faces an especially awkward version of the branching problem. In the usual uploading stories, the scanner erases the brain. But it needn't: the information could be read off without any erasure, then copied to a computer and used to create a person there as in the original story. Transhumanists call this "nondestructive

13 See e.g. Shoemaker, "Personal Identity," 85; Parfit, *Reasons and Persons*, 207.

14 For a good discussion of this problem, see Harold Noonan, *Personal Identity*, 2nd ed. (London: Routledge, 2003), 127–39.

uploading." The result would be two people—one human, one electronic—each psychologically continuous with you. According to the nonbranching proposal, neither would be you, as this would be a case of branching. And there is no one else after the transfer that you could be. You would cease to exist. Nondestructive uploading may appear harmless, but would in fact be fatal.[15]

Transhumanists may reply that you *could* survive branching in this case: if the procedure leaves your brain intact, you carry on as usual, and the electronic person created in the computer is someone new. But although that's surely right, it only creates another mystery: why is it possible to survive "asymmetric" but not "symmetric" branching? The idea behind uploading is that copying the psychological information from your brain to a computer would move you to the computer. Why should it do so only if that information is erased from your brain? Why is that act of destruction necessary to send you on your way?

It's easy to see that the same mysteries arise in cases not involving uploading: we need only imagine a variant of the transporter that produces two copies the Captain on board the ship, or one that scans him without dispersing his atoms.[16]

The Duplication Problem

A second worry about the personal-identity assumption arises from the fact that there is a difference between any concrete object and a copy or replica of that object, no matter how exact. I don't mean a qualitative difference. A replica of the Rosetta stone might be entirely indistinguishable from the original, right down to its subatomic structure, but still the replica would be one thing and the original another. One would have been created by geological processes millions of years ago

15 David Chalmers hesitantly accepts this: "The Singularity: A Philosophical Analysis," *Journal of Consciousness Studies* 17, nos. 9–10 (2010): 55.

16 Parfit, *Reasons and Persons*, 200–201.

and carved by Egyptians in the second century BCE. The other would have been created only today by the Martians.

The same goes for people: there's a difference between a person and a replica of that person. There could be a replica of Ludwig Wittgenstein as he was at any moment during his life. It may resemble him in all intrinsic respects, or it may be only a psychological replica, with all his intrinsic mental properties but physically different. The AI assumption implies that we could create a psychological replica of Wittgenstein by programming a computer in the right way. And the personal-identity assumption implies that by doing so we could upload Wittgenstein himself.

Now imagine that the Martians (who have all the technology that we lack) visited the earth shortly before Wittgenstein's death and made a detailed scan of his brain. The British Wittgenstein Society acquires a copy of the scan, and proposes to use it to create a psychological replica of him as he was then, so that they can put to the replica all the questions about Wittgenstein's philosophy that have accumulated in the intervening decades. (They have a long list.) The replica would be able to answer their questions just as well as the original could. They don't want to bring back Wittgenstein himself because the interrogation will be stressful and they think he deserves his rest. The Austrian Wittgenstein Society, however, has no such scruples, and wants to use its own copy of the scan to bring back the great man himself.

If human beings can be uploaded, both endeavors should be possible: Wittgenstein himself can exist in a computer, and so can another person psychologically just like him. But what would the two societies have to do differently so that the Austrians got the original and the British got a replica? There appears to be nothing they *could* do differently. To create a psychological replica of Wittgenstein as he was at the time of the scan, the British would have to program the information into a computer so as to create a person with the mental properties that Wittgenstein had then. The Austrians would do precisely the same—and according to the personal-identity assumption, that would

recreate Wittgenstein himself. (To avoid complications to do with branching, suppose that only one society actually carries out its plan.)

It would follow that there is no difference between bringing Wittgenstein back to life and creating a mere replica of him. There is, accordingly, no difference between your being uploaded into a computer and the creation there of a new person psychologically just like you. This conflicts not only with the indisputable fact that there *is* a difference between an original object and a copy, but also with the personal-identity assumption: if you yourself, as opposed to a mere replica of you, could exist in a computer, there must be a difference between these two states of affairs.

Or maybe the problem is not that there would be no difference between originals and copies, but that the difference would be nothing like what we thought it was. Again, the personal-identity assumption is based on the view that our persistence consists in psychological continuity with any reliable cause (and perhaps no branching). It follows that any being sufficiently like Wittgenstein in his mental properties must be Wittgenstein himself, as long as the resemblance is not accidental (and there's no branching). In their attempt to do him the kindness of creating only a replica, the British would inadvertently resurrect the poor man from his quiet grave in Cambridge. That's like saying that if the owners of a Las Vegas hotel set out to build a replica of the Colossus of Rhodes on the basis of newly discovered blueprints, and the resemblance were near enough, they would have on their hands not a modern replica but the original statue—a historic artifact cast thousands of years ago in the foundries of ancient Greece. The modern-day Greeks could reasonably ask for it to be given back.

Material Things and Material Continuity

I have argued that the personal-identity assumption has troubling implications about branching and about the difference between originals and replicas. These are good reasons to doubt whether uploading

is metaphysically possible. But I haven't explained *why* it isn't. Pointing out that uploading would have absurd consequences may show that it could never happen, but it won't tell us why it couldn't. What is it about the procedure that prevents it from moving us to a computer?

I think the reason we cannot be uploaded is that we're material things. We're made entirely of matter. And a material thing cannot continue existing without some sort of material continuity.[17] It must continue to be made up of some of the matter that made it up previously. It can change all of its matter if this happens gradually enough, but it can't change all its matter at once. It follows that you can't move a material thing from one place to another merely by transferring information. You can't send a human being as a message by telegraph (despite the joke in *Alice in Wonderland*)—or as a text message or an email attachment.

And there is no material continuity in uploading. No matter moves up the wires from a human organism to a computer. (If something seems to move, consider that the information could as well be written down in a letter or dictated over the phone.) You can't upload a human being for the same reason that you can't upload a tree or a brick.

We can make the point more vivid by thinking about what sort of material things we might be. We might be biological organisms.[18] (The appearance that we're material things is arguably the appearance that we're organisms. If you examine yourself in the mirror, you see an organism. It seems the same size as you—no bigger or smaller. We appear to have all the physical properties of human organisms, and the same behavior. They certainly don't appear to be *other* things

17 For a powerful defense of this claim, see Peter van Inwagen, "Materialism and the Psychological-Continuity View of Personal Identity," in *Philosophical Perspectives 11: Mind, Causation, and World*, ed. James Tomberlin (Malden, MA: Blackwell, 1997). Corabi and Schneider say that we can't be uploaded because this would involve a gap in our existence and this is impossible, but their account of why it's impossible is obscure: Joe Corabi and Susan Schneider, "The Metaphysics of Uploading," *Journal of Consciousness Studies* 19, nos. 7–8 (2012): 26–44. Temporal-parts theorists can avoid the problem: see Olson, "The Central Dogma of Transhumanism."

18 I defend this view in chapter 2 of *What Are We?*

than ourselves.) But you can't move a biological organism—a human animal or a dog or a tree—to a computer by scanning it and uploading the information thereby gathered. Scanning may leave the organism unharmed. Or it may damage it, perhaps fatally. It may even completely destroy the organism by dispersing its atoms, as the *Star Trek* transporter does. But no matter what form the scan takes, the organism stays behind. Our being organisms would make uploading metaphysically impossible.

We might of course be material things other than organisms. We might be brains, literally made up entirely of soft, pinkish tissue and located within the skull.[19] But scanning your brain can't remove it from your head. The organ may remain unchanged in the scanning process, or it may be damaged, or even completely destroyed by having its atoms dispersed; but it doesn't move to a computer. A brain can no more be uploaded than a foot can.

The same goes any other material thing. If you scan it and transfer the information thereby gathered to a computer, whether electronically, in writing, or orally, the material thing stays where it is. Our being uploadable would rule out our being material things of any sort. Each of us would have a property that no material thing could have: the capacity to be sent as a text message. The reason we can't be uploaded is that we're material things and it's metaphysically impossible to move a material thing without moving matter.[20]

Some metaphysicians say that certain material things *can* move by a mere transfer of information, even if organisms and brains can't. Specifically, human beings can. The thought is that human beings are

19 Derek Parfit, "We Are Not Human Beings," *Philosophy* 87 (2012), 5–28; Hud Hudson, "I Am Not an Animal!," in *Persons: Human and Divine*, ed. Peter van Inwagen and Dean Zimmerman (Oxford: Oxford University Press, 2007), 216–36. For discussion and further references see Olson, *What Are We?*, 76–98.

20 This does not rule out our becoming electronic people by gradually replacing our organic parts with bits of computer hardware till none remain. I lack the space to discuss this view. But I know of no one who thinks that we could survive gradual transformation into electronic people but not uploading.

material things "constituted by" organisms.[21] We're made of the same matter as human organisms, and are physically indistinguishable from them. But although *they* need material continuity to survive, we don't. In uploading, a person is constituted first by a human organism and then by a computer.[22]

There is a large debate about whether "constitution" is even possible.[23] But in the current context this is a distraction. Our being constituted by organisms would do nothing to show how material things could survive without material continuity. If you want to explain how a material thing can be sent as a text message, it's no help claiming that it shares its matter with another thing that *can't* be sent as a text message. Nor, come to that, does the proposal suggest any solution to the branching or duplication problems.

The Pattern View

Those are my reasons for doubting the possibility of uploading. The rest of this essay will consider replies. The most obvious is this: if material things can't be uploaded, could we not be *im*material things—things not made of matter?

This would have a dramatic implication: that it's metaphysically impossible for any material thing to think or be conscious. If human beings are immaterial, then presumably all beings with mental properties must be immaterial. If it were ever possible for any conscious being to be made of matter, *we*'d be made of matter. We *seem* to be material things. When you look at yourself or any other human being, you see nothing but flesh and bone. We're as material as any conscious being

21 Shoemaker, "Personal Identity," 108–14; Lynne Rudder Baker, "Death and the Afterlife," in *Oxford Handbook for the Philosophy of Religion*, ed. William Wainwright (Oxford: Oxford University Press, 2005): 366–39.

22 Lynne Rudder Baker, *Persons and Bodies: A Constitution View* (Cambridge: Cambridge University Press, 2000): 109.

23 For a summary with references, see Olson, *What Are We?*, 48–75.

could ever be. If we're *not* material, despite every appearance to the contrary, that can only be because all conscious beings must be immaterial.

Or again: if human beings could be uploaded by having their psychological information programmed into a computer, then any thinking or conscious being could be. It could hardly be that some thinking beings are uploadable and others are not. What could account for this difference? And what reason could we have to suppose that we belong to the uploadable kind? Yet as we've seen, no material thing is uploadable: that was the reason for supposing that we must be immaterial. Any thinking or conscious being must therefore have a property that no material thing could have, namely being uploadable. It follows that no material thing could think or be conscious.

So there are thinking or conscious things and there are material things, but nothing can be both. The property of being conscious and the property of being made of matter are metaphysically incompatible. What we ordinarily take to be a conscious human being made of matter is really two things: a conscious, immaterial thing (traditionally called a "soul") and an unconscious, material thing. This is the view known as substance dualism. Transhumanists don't want to be substance dualists. Unless a material thing can somehow be sent as a text message, however, they can't avoid it.[24] If we can be uploaded but material things cannot be, we must inevitably conclude that we're not material things. And whatever is not a material thing is an immaterial thing.

But simply denying that we're immaterial things is not enough to defend the possibility of uploading. How would one go about transferring an immaterial thing—a soul—from a human organism to a computer? The suggestion has to be that we're immaterial things of a sort that could be uploaded.

24 Or they could deny the reality of the material world altogether and embrace idealism. In that case all bets are off.

Transhumanists often say that a person is not a material thing but rather a sort of pattern or bundle of information. Bostrom claims that we might one day "live as information patterns on vast super-fast computer networks."[25] Ray Kurzweil (another enthusiastic transhumanist) says that because living organisms constantly exchange matter with their surroundings, "all that persists is the pattern of organization of that stuff . . . like the pattern that water makes in a stream as it rushes past the rocks in its path. . . . Perhaps, therefore, we should say that I am a pattern of matter and energy that persists over time."[26] This thought is not confined to transhumanists: Daniel Dennett suggests that "what you are is that organization of information that has structured your body's control system."[27] We say that the same organization or pattern is present in the A-team during the first half of the football match and in the B-team during the second half. Might it not be present first in a biological organism and then in a computer? If so, the scanning-and-uploading procedure that transhumanists imagine would bring it about. If human beings are such patterns, that might make us uploadable.

The proposal must be that a human being—the author of this essay, for instance—is literally a pattern. It can't be merely that mental states or events are information patterns, or that to think is to exemplify a certain sort of pattern. That would do nothing to explain how someone could move from an organism to a computer. (In fact it's incompatible with our being organisms.) It must be a view about the metaphysical nature of thinking beings.

As I see it, the attraction of the "pattern view" is due entirely to its vagueness. It's rarely stated in enough detail to make clear what would

25 Nick Bostrom et al., "Transhumanist FAQ, 3.0," http://humanityplus.org/philosophy/transhumanist-faq/, accessed August 9, 2016.

26 Ray Kurzweil, *The Singularity is Near: When Humans Transcend Biology* (London: Duckworth, 2006), 383.

27 Dennett, *Consciousness Explained* (Boston: Little, Brown, 1991), 430. For the suggestion that a computer program could be a person, see Dennett, "Where Am I?," in *Brainstorms* (Cambridge, MA: MIT Press, 1978), 310–33.

actually follow from it. Once this is done, it becomes impossible to take seriously.

What is a pattern? Not a material thing—that was the whole point of the proposal. It must be something that can in some sense be present in different material things: first in an organism and then in a computer, for example. Presumably a pattern is a universal rather than a particular—a type rather than a token, like a novel rather than a particular hard copy of a novel. The proposal must be that each of us stands to a particular human organism as *Brighton Rock* stands to the old paperback on my shelf, except that each of us exists in only one copy. The person, like the novel, is not something we can see or touch. It's abstract and intangible. It's present in the physical world only in the way that justice is present when the legal system is working properly, or that the number eleven is present when all the players are on the field.

That still leaves many questions unanswered, but it's enough to start with. To my knowledge, no professional metaphysician—no one whose job it is to think systematically about the fundamental nature of concrete and abstract objects—has ever held such a view. Here are just two obvious objections.

First, think about which pattern a human being might be. Given the length of my life, I'd have to be a pattern sufficiently general and undemanding as to be shared by a drooling infant, a moody teenager, a middle-aged academic, and a senile old man. This human organism exemplifies many different patterns as it changes. But there would have to be a single pattern that is always present in it, and which is compatible with any course my life may take in the future.

It may be that patterns are abundant, so that there's bound to be *some* highly variegated pattern that this organism exemplifies throughout its life. But in that case there are millions of other patterns that it exemplifies for shorter periods: years, days, hours, seconds. What could make it the case that of all these patterns, just one of them was me? Why one rather than any of the others? The pattern that I am would of course have to be conscious and think my thoughts: if I know anything about

myself, I know that I'm conscious and thinking. But how could just one of the patterns be conscious? And why should it be one that's variable enough to be present in this organism from cradle to grave? If any pattern of information could be conscious, wouldn't a large number of those present in this organism be so, including some that are present for only a day or an hour? That would leave me in the absurd predicament of wondering which of these many thinking patterns was me, and whether I existed last year or came into being only yesterday. I can't see how anything that can legitimately be called a pattern of information could both last throughout my life and be the only thinker of my thoughts.

The second problem arises when we recall the "British" and the "Austrian" Wittgenstein from my fanciful story. (Imagine now that they exist simultaneously.) The pattern view implies that the original Wittgenstein is a single pattern that they both exemplify. (Ignore the problem of which pattern it is.) So the two latter-day Wittgensteins are in one sense the same philosopher and in another sense not: they're two tokens of the same type, like two hard copies of *Brighton Rock*. Now suppose that the British Wittgenstein is awake and thinking hard at a time when the Austrian Wittgenstein is asleep and unconscious. Is Wittgenstein awake or asleep?

We might say that he's awake in the sense of having a concrete instance that's awake, and asleep in the sense of having another instance that's asleep, just as *Brighton Rock* is both torn, insofar as my copy is torn, and intact, insofar as yours is undamaged. But the pattern view implies that strictly speaking, Wittgenstein himself neither wakes nor sleeps, just as *Brighton Rock* itself is neither torn nor intact. He doesn't really think. Only his particular instances do these things—the concrete, material things that exemplify the pattern. But of course he does think: he is, after all, a philosopher. It follows that Wittgenstein is not a pattern but at best something that exemplifies certain patterns. And because we ourselves are thinking, conscious beings, we're not patterns either.

Transhumanists cannot avoid the metaphysical problems facing their view simply by denying that we're material things.[28]

Parfitian Transhumanism

Transhumanists need to say what sort of thing could be conscious, intelligent, and transferrable from an organism to a computer. And they need to solve the branching and duplication problems. They have their work cut out for them.

Suppose we can't be uploaded. The most we could get by scanning your brain and programming the information extracted from it into a computer is a psychological replica of you. Transhumanists might be willing to concede all this. They could say that although uploading could not give us numerical identity between the scanned human person and the replica in the computer—it couldn't make them one and the same—it could give us what matters practically in identity. Even if an electronic person could not literally be you, it might be just as good, as far as anyone's interests are concerned, as if she were. It's not a difference that we have any reason to care about.[29]

The thought, advocated by Derek Parfit and others, is that your interest in continuing to exist is not an interest in there being a future person who *is* you—that is, numerically identical with you—but in there being a future person bearing some other relation to you. Parfit thought this relation was psychological continuity, broadly construed so as not to require continuous physical realization of mental properties (so that it could hold in cases of *Star Trek* teleportation) and without any nonbranching restriction (so that it can hold between you

28 For further objections to the "pattern view," see Olson, *What Are We?*, 145–49. Another species of dualism that transhumanists might explore is the Humean view that each of us is a "bundle of perceptions." But that's a large topic and I can't discuss it here.

29 This may be the view of Eric Steinhart in *Your Digital Afterlives* (London: Palgrave Macmillan, 2014), 62–68, though there is much in his account that I don't understand.

and two future people).[30] His view was that I have no selfish or prudential reason to care whether I myself exist tomorrow. What I have reason to care about is only whether someone exists then who is psychologically continuous with me as I am now.

In all real cases, any future being who is psychologically continuous with me really is me. But not in all possible cases. I might undergo branching, so that neither resulting person is me and I cease to exist. But because they'd both be psychologically continuous with me, this would be just as good for me as if I actually survived. And so would destroying me and creating a psychological duplicate of me in a computer.

Parfit also thought that whether I have a selfish or prudential interest in someone's future welfare depends not on whether that person is me, but only on whether he's psychologically continuous with me. Someone's being psychologically continuous, tomorrow, with me as I am now gives me the same reason to care about his welfare then that I have to care about my own welfare. If he has to spend the next week marking undergraduate essays, I have the same reason for dread as when I myself have to spend the next week marking essays.

The proposal, then, is that the existence of an electronic person in a computer could give me everything I have reason to want in wanting to continue living. I could have the same reason to care about his welfare as I have to care about my own—a reason I'd have even if I were completely selfish and would not lift a finger to save my own mother from unbearable agony. Whether he would literally be me is of merely theoretical interest. All that matters practically is whether he would be psychologically continuous with me. Transhumanists could then dispense with the personal-identity assumption. Even if we can't enter the promised electronic realm ourselves, our "Parfitian successors" can

30 Parfit, *Reasons and Persons*, 262. More precisely, it's psychological continuity combined with a degree of "connectedness"—roughly psychological similarity. The difference is unimportant for my purposes here: in all the cases discussed here there is both continuity and connectedness.

dwell there, and for us that's just as good. We might call this *Parfitian transhumanism.*

The trouble with this proposal is that the idea it's based on looks false. The existence of a future person psychologically continuous with me does not appear to have the same practical consequences as my surviving. Someone's being psychologically continuous with me does not by itself seem to give me any selfish reason to care about his welfare. Consider again the case of "nondestructive uploading." Suppose I'm kidnapped, and the Martians are going to scan my brain and use the information thereby gathered to create an electronic person psychologically just like me. Because the scanning is harmless, the process will leave behind a human being exactly like me, and materially continuous with me to boot. Nearly everyone would say that he *is* me. Suppose this is right. Yet both people will be psychologically continuous with me as I am now. One will then be brutally tortured. Otherwise they'll be treated identically: perhaps both will be suddenly and painlessly killed, one after a week of agony and the other after a week spent pleasantly. Though the torture will be the same no matter what I do, my captors allow me to choose, before the uploading takes place, which one gets it: me or the electronic person.

If uploading preserved what matters practically, there would be no reason for me to prefer one alternative over the other. I'd have the same selfish reason to care about the electronic person's welfare as I have to care about my own. I may as well toss a coin. But that's not how it seems. I don't know about you, but I would far rather have the electronic person tortured. I suspect, in fact, that if I were entirely selfish I'd be completely indifferent about his welfare. My only concern would be for myself.

Or imagine that the Martians learn how to scan people's brains without their noticing, making kidnapping unnecessary. They then upload the information from the scan into a computer, creating an electronic person psychologically identical to the original as in the previous story. The electronic person is then tortured. Suppose the

Martians have been active in my neighborhood and there is a real chance that they will scan my brain tonight as I sleep and torture the resulting electronic person. If uploading preserved what matters prudentially, I ought to be just as worried about this as I'd be if I thought there was a real chance that I myself was going to be tortured. But anyone would find the second case far more frightening.

Of course, we may be wrong about what matters in these cases—that is, about what we have a prudential reason to care about. Maybe it would be just as irrational for me to care about my own welfare but not about that of the electronic person created by the Martians as it would be for me to care what happens to me on Tuesdays and Thursdays but not what happens to me on other days. But that would make Parfitian transhumanism hard to believe.

Final Remarks

It may, perhaps, be possible to create purely electronic people mentally superior to ourselves and free from human limitations. We may even have reason to do so: these beings might be able to carry out our projects far better than we ever could. They may be able to survive the changes to the climate that are likely to devastate living things. But even if all this is true, it doesn't look as if we human beings could become purely electronic.

For valuable comments on earlier versions of this essay I am grateful to Karolina Hübner, Marcy Lascano, Yair Levy, Luis López-Farjeat, Eli Pitcovski, Eric Schliesser, Katherine Withy, and Karsten Witt.