1. The Central Dogma Of Transhumanism

ERIC T. OLSON

1. The Central Dogma

Transhumanism is a movement aimed at enhancing and lengthening our lives by means of futuristic technology. The name derives from the ultimate goal of freeing us from the limitations imposed by our humanity. Human beings are subject to many ills: disability, exhaustion, hunger, injury, disease, ageing, and death, among others. They set a limit to the length and quality of our lives. 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 in the biological sense–better yet, if we could cease to be biological at all–these limitations could be overcome. An inorganic person would not be subject to exhaustion, disease, ageing, or death. The length and quality of her life could be extended more or less indefinitely. So it would be a great benefit, transhumanists say, if we could make ourselves inorganic.

They hope to achieve this by a process they call "uploading." The information in your brain is to be transferred to an electronic digital computer. The process does not merely store the information on the computer, as when you upload a letter of reference to a distant server, but uses it to create a person there: a being psychologically just like you, or at any rate a great deal like you. This person may be psychologically human, but not biologically. He or she would not be made of flesh and blood.

The aim is not merely to create new people in computers, but for *us* to move from our human bodies to the digital realm. The thinking is that the person created by the uploading process would be psychologically continuous with you: her mental properties would resemble and be caused by yours in much the same way that the mental properties you have now resemble and are caused by those you had yesterday. Given the widely held assumption that this is what it *is* for a person to continue existing–that personal identity over time consists in psychological continuity–the person in the computer would be you.

And once you are in or on a computer, you needn't worry about disease or injury or ageing or death. If the computer hardware that houses you is damaged, you need only move electronically to another piece of hardware. Travel would be as easy as emailing. You would not need food or shelter or furniture. The limitations imposed by human biology, or indeed any biology, would be a thing of the past. Your intelligence, patience, capacity for pleasure, and physical strength and stamina (if you are given a robotic body) could be enhanced indefinitely.

These hopes are founded on the extravagant assumption that the technology of tomorrow will literally make it possible to transfer a person from a human organism to a computer. Call this the *central dogma* of transhumanism. (The name is not meant to be pejorative; think of the central dogma of molecular biology.) The leading transhumanist Nick Bostrom puts it like this:

> 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). (Bostrom 2001)

Bostrom and others are confident that that we *could* "scan the synaptic matrix of a human brain and simulate it on a computer," and thus that such "migration" is possible.

The central dogma is of more than merely theoretical importance. If it really were possible for us to move from our human bodies to electronic computers, subject only to limitations of technology, it would mean that we are not doomed to wither and die. We are at least potentially immortal.

The central dogma raises many large questions. One is whether a "post-human" life would be as attractive and worthwhile as transhumanists imagine. Another is whether any of this is likely ever to happen. This paper is about the worries Bostrom puts in parentheses: whether it is metaphysically possible.

2. The Dogma's Presuppositions

The central dogma presupposes three contentious claims. The first is that there can be genuine artificial intelligence: it is possible for a computer not only to simulate intelligence and consciousness, but actually to *be* intelligent and conscious. More precisely, a computer could have the mental properties that you and I have. This will of course include those that make something a person, as opposed to a being with mental properties that fall short of those regquired fot personhood in the way that, for instance, those of dogs do: such properties as self-consciousness. So it must be possible to

create a person just by programming a computer in the right way (and perhaps also providing appropriate connections to the environment). In other words, an electronic computer could be a person. Or perhaps we should say not that a computer could actually *be* a person, or be conscious and intelligent, but rather, more vaguely, that it could "realize" or "implement" a person or a conscious and intelligent being. (I will return to this point in a moment.) Call such a being a *computer person*. So the first presupposition of the central dogma is that there could be a computer person. This is what Bostrom means by "the assumption about the nature of consciousness."[1] I will call it the *AI assumption*.

The second presupposition is that you and I could *become* computer people. This is what Bostrom means by "the assumption about personal identity." It presupposes the AI assumption but does not follow from it. If I could become a computer person, then computer people must be possible; but the mere possibility of computer people does not imply that we ourselves could become such people. By analogy, it may be that there could be gods–conscious, intelligent beings who are immaterial and supernatural–even if it is metaphysically impossible for us to become gods.

In this regard the central dogma is like the doctrine of the resurrection of the dead: the claim that when we die and our physical remains decay, we do not perish, but continue existing in a conscious state in the next world–a place spatially or temporally unrelated to this one. This presupposes that there *is* a next world, some of whose inhabitants are people psychologically like us. But the mere existence of such a place would not make it possible for someone to get there from here. How could it be that I am totally destroyed in the grave, yet at the same time continue to exist with my psychology intact in the next world? That is the metaphysical obstacle to resurrection (van Inwagen 1978, Olson 2015). Transhumanism faces an analogous obstacle: how could it be that I am totally destroyed in the grave, yet continue to exist with my psychology intact in a computer? How is the "uploading" procedure supposed to bring this about?

The personal-identity assumption has an immediate and important implication, namely that uploading would not transform the computer itself–the physical object made of metal and silicon and plastic–from a nonperson to a person. This is because (according to the assumption) the person who ends up in the computer was previously in a human organism. She was not previously in the computer as a nonperson. It is the human person

[1] In calling it an assumption about the nature of consciousness rather than about the nature of the mental in general, Bostrom is presumably taking it to be uncontroversial that computers could have mental properties that do not require consciousness. This is doubtful, but I won't press it.

who becomes a computer person, rather than the previously unintelligent computer becoming a computer person. This implies that no computer could ever be a person itself. If a computer *could* ever be a person, or be conscious and intelligent, it could be made so by uploading–that is, by programming it in the right way. But in that case uploading would create *two* people or conscious beings: the former human person and the former unintelligent computer. The two computer people would be psychologically indistinguishable. Both would seem to remember my embodied past, one correctly and one falsely. How could either of them ever know which one he is? I take that to be absurd. So the personal-identity assumption entails that no computer could be conscious or intelligent. At best a conscious, intelligent being might "inhabit" or "be implemented on" a computer. (I will return to the question of what this "inhabiting" relation might be.)

The third presupposition of the central dogma is that it is possible for technology to advance to the point where we could actually do these things. This presupposes the first two claims, but does not follow from them. Even if uploading a human person into a computer is metaphysically possible, it may remain beyond any possible human capability. We might compare it with the task of creating a perfect physical duplicate of a human being. This is metaphysically possible: God could do it. But it's doubtful (to put it mildly) whether it could ever be possible for *us* to do it. Uploading might be like that.

I see no reason to feel hopeful about this third assumption, even if the others are true. But my interest is in the metaphysical assumptions, especially the one about personal identity.

3. The Branching Problem

Suppose for the sake of argument that the AI assumption is true: it is possible to make a digital computer into a person–or rather, to get it to "implement" or "realize" a person–by programming it in the right way. Even so, could a human person literally move to a computer? Transhumanists have had little to say about this. Some have defended the AI assumption at length (Chalmers 2010), but once they have established to their satisfaction that a person could exist in or on a computer, they have seen little reason to doubt whether we ourselves could do so. I think there are strong reasons for doubting it.

Here is one obvious worry. If someone could be uploaded into a computer, then someone could be uploaded into two computers. That is, the relevant information could be read off the human brain and copied simultaneously to two separate and independent pieces of computer hardware in just the way that transhumanists envisage its being copied to one. The

result would be two computer people, each psychologically just like the original human person. Each would have got his or her mental properties from the original person in the same way. So nothing could explain why one but not the other was the original person. More strongly, it seems that nothing could make it the case that one but not the other was the original person. If one were the original person, both would be. But they couldn't both be. There are two computer people in the story and only one human person, and one thing cannot be two things. If the original person and the first computer person are one, and the original person and the second computer person are one, then the first computer person and the second computer person would have to be one. (If $x=y$ and $x=z$, then $y=z$.) But they're not.

It appears to follow that a person could not move from a human body to a computer in the "double-upload" case. And if it's not possible in the double-upload case, it could hardly be possible in the "single-upload" case commonly imagined, because the same thing happens in both: the same information from the person's brain is transferred to a computer in the same way. So no amount of uploading is sufficient to make a human person into a computer person, contrary to the personal-identity assumption. Call this the *branching problem*.

The branching problem is familiar to anyone acquainted with philosophical discussions of personal identity. The reason is that it arises on almost any version of the psychological-continuity view–any view to the effect that an earlier person is the same as a later person just if the later person is in some way psychologically continuous, at the later time, with the earlier person as she is at the earlier time. (Psychological continuity is defined in terms of causal dependence of later mental states on earlier ones; for details see Shoemaker 1984: 90.) The most popular accounts of personal identity over time are of this sort. And it's clear that the personal-identity assumption implicit in the central dogma of transhumanism presupposes a psychological-continuity view: the reason why transhumanists think you could become a computer person is that they think a computer person could be psychologically continuous with you.

The most commonly proposed solution to the branching problem is to deny that someone's being psychologically continuous with you in the future suffices for you to survive. What suffices is, rather, "non-branching" psychological continuity. A later person is you just if she is psychologically continuous with you *and* there is no branching (e.g. Shoemaker 1984: 85; Parfit 1984: 207). The implication in the "uploading" case would be that as long as the psychological information from your brain is uploaded only once, the resulting person is you; but if it were simultaneously uploaded

more than once, none of the resulting people would be you. Each would be a newly created person mistakenly convinced that she was you and with false memories of your life, including the belief that she had been alive for many years. It is metaphysically possible for a person to move to a computer by "single upload" but not by "double upload."

The obvious and well-known objection to this is that non-branching requirements are arbitrary and unprincipled. The claim that you could survive single but not double uploading is surprising. And the proposal does nothing to explain *why* the occurrence of a second uploading procedure would prevent the first such procedure from moving you to a computer. Why should an event that would normally suffice to preserve your existence destroy you if accompanied by another instance of the same procedure–something that has no causal effect on the first event? What is it about the second upload that destroys you? The only answer seems to be that surviving a double upload would lead to a logical contradiction: to one thing's being numerically identical to two things. But that can't be the whole story. It cannot be merely the laws of logic that prevent us from surviving double uploading.

The current proposal faces a particularly awkward version of the branching problem. In the usual uploading stories, the brain is conveniently erased in the scanning process. But this need not be so: the relevant information could be "read off" without doing you any damage, then copied to a computer and used to create a person there exactly as before. For you it might be like having an MRI scan. Tranhumanists call this "nondestructive uploading." The result would be two people–a human person and a computer person–each psychologically continuous with you. But according to the non-branching proposal, neither would be you, as this would be a case in which two people come to be simultaneously psychologically continuous with you. And there is no other being after the transfer that you could be. It follows that you would cease to exist: nondestructive uploading would be fatal.

If this isn't already troubling enough, it raises an awkward epistemic problem. For all I know, the Martians (who have all the advanced technology that we lack) could be scanning my brain right now and copying the information to a computer, thereby creating a person psychologically continuous with me. It follows from the non-branching requirement that I could cease to exist at any moment, mid-sentence, without the slightest disruption of my mental life or physical functioning, and be instantly replaced by a new person with false memories of my life. No one would be any the wiser. It is hard to take this seriously.

Transhumanists are likely to respond by saying that it *is* possible to survive branching in this case: if the uploading procedure leaves your brain intact, you continue existing as you are, and the computer person thereby created is someone new. That, of course, sounds right. But this new proposal adds a second arbitrary and unprincipled feature to the first one. Why could someone survive "asymmetric" but not "symmetric" branching? Why, in other words, would transferring the information from your brain to a computer be "person-preserving" (as psychological-continuity theorists like to say) if, but only if, that information is gathered in a destructive way? And why, after the uploading, would you be the person with your body and not the person in the computer? The obvious answer is that you would survive as the person with your body because he or she would be materially or biologically continuous with you, and the person in the computer would not be. But the possibility of surviving ordinary, "single" uploading would imply that we can survive without material or biological continuity. Why is material continuity suddenly relevant here? The only answer would seem to be that appealing to it can avoid implausible consequences. But again, what enables me to survive asymmetric but not symmetric branching cannot be the fact that it would be implausible to suppose otherwise.

4. The Duplication Problem

Here is a second and less familiar worry about the personal-identity assumption. There has to be a difference between me and someone psychologically just like me. Someone could be a perfect psychological duplicate of me as I am at some particular time–now, say–without being me. There is a difference between a particular person and a copy or replica of that person, no matter how exact, just as there is a difference between the original Rosetta stone and a replica of it created today, no matter how exact. I don't mean a qualitative difference. A replica of the Rosetta stone might be completely indistinguishable from the original, right down to its finest atomic structure. Still, the replica would be one thing and the original would be another. The original would have been made by hand in the second century BC; the replica would have been made only today by the Martians.

So there could be a replica of Wittgenstein as he was at any moment during his life. It might resemble Wittgenstein in all intrinsic respects–a flesh-and-blood being, atom-for-atom identical to him–or it may be merely a psychological replica, with all his intrinsic mental properties but physically different. The AI assumption implies that we could create such a replica simply by programming the right sort of computer in the right way, if only we had in our possession the psychological information realized

in Wittgenstein's brain at the appropriate time. And the personal-identity assumption implies that this knowledge would enable us to upload Wittgenstein himself into a computer, abruptly resurrecting him from his quiet grave in Cambridge.

Imagine, then, that the British Wittgenstein Society somehow get access to a detailed scan of Wittgenstein's brain made shortly before his death. They propose to use it to create a psychological replica of him as he was then, so that they can put to him all the questions about his work that have accumulated in the intervening decades. (They have a long list.) A psychological replica of the man would be just as willing and able to do this as Wittgenstein himself would be. But they want a replica and not the original because they fear the interrogation will be traumatic, and they feel that Wittgenstein has suffered enough for philosophy already. The Austrian Wittgenstein Society, however, have no such scruples. They have their own copy of the scan, and want to use it to bring the great man himself back to life in order to attract foreign visitors.

If the central dogma is true, both projects are possible. The question is, what would the two societies have to do differently so that the Austrians got the original Wittgenstein and the British got a replica? It looks as if there is 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 copy the psychological information from the scan to a computer in such a way as to create a conscious, intelligent person with just the intrinsic mental properties that Wittgenstein had at a certain time in 1951. The Austrians would of course do precisely the same thing. And according to the personal-identity assumption, that would suffice to upload Wittgenstein himself into the computer. It would follow that there was no difference between bringing Wittgenstein himself back to life and creating a brand-new replica of him. Likewise, there would be no difference between your being uploaded into a computer and someone else's being newly created there. This conflicts not only with the indisputable fact that there *is* a difference between an original object and a copy, but also with the central dogma, which says that you yourself, and not merely a copy of you, could exist in a computer. Call it the *duplication problem*.

5. Why the Problems are Superficial

The branching and duplication problems are serious, and transhumanists have had little to say about them. But I don't think the problems go very deep. If uploading really is metaphysically impossible, it cannot be for these reasons–because it has absurd consequences about personal identity

over time and about the difference between originals and duplicates. These consequences are symptoms of a deeper, underlying problem.

We can see that the branching and duplication problems do not strike at the heart of the central dogma by noting that they apply equally to claims that do not involve uploading. One is that a person could travel by *Star Trek* teleportation. Suppose the teleporter works like this. When the Captain has had enough adventures on the alien planet, the teleporter "scans" him, thereby dispersing his atoms. The information gathered in the scan is then sent to the ship, where it is used to assemble new atoms precisely as the Captain's were arranged when he said, "Beam me up!" The result is someone both physically and mentally just like the Captain. And it's part of the story that the man who materializes on board the ship *is* the Captain.

If the man appearing on the ship really could be the Captain, the branching problem would apply just as it does in the case of uploading: the teleporter could produce two beings like the Captain instead of one. And if the man who appears in single teleportation would be the Captain, both men who appeared in double teleportation would be, with the impossible result that one thing is numerically identical to two things. Avoiding this problem by introducing a non-branching clause would imply that if I were scanned in a way that did not disperse my atoms and the information thereby gathered were used to assemble an exact duplicate, that would be the end of me, as it would be a case of branching.

Likewise, the information gathered in the scan could be used either to create a replica of the Captain or to recreate the Captain himself; yet the procedure for doing both these things would be exactly the same. It would seem to follow that there was no difference between a person and a replica of that person.

Another view with similar implications is Shoemaker's claim that a person could move from one organism to another by what he calls "brain-state transfer."[2] He imagines a machine that scans your brain just as in the uploading story, thereby recording all the relevant information realized in it and erasing its contents in the process. This information is then transferred not to a computer, but to another human organism with a "blank" brain, resulting in someone psychologically just like you (or as much like you as the new organism's physical properties allow). Shoemaker claims that because this being would be psychologically continuous with you, he or she would *be* you–as long as the machine copies your brain states only once and your original brain is erased. It's easy to see that the same worries about branching and duplication apply here as well.

[2] Shoemaker 1984: 108-111. I don't know whether any other philosopher has ever shared this view.

These views have nothing to do with uploading. They could be true even if the central dogma were false and uploading were impossible. Whatever makes teleportation and brain-state transfer impossible, if indeed they are, must be something independent of the AI and personal-identity assumptions.

Not only are the branching and duplication problems not peculiar to uploading, but there may be species of uploading that avoid the problems. Suppose the uploading process took place bit by bit rather than all at once. A small portion of your brain is scanned, and its functions, or at any rate those that are relevant to your mental properties, are duplicated in a computer. (If a computer can duplicate the functions of your entire brain, it can duplicate the functions of part of it.) The neurons communicating with the scanned brain part are then connected to the computer by radio links, and the scanned brain part itself is destroyed or disabled. The result is that your mental activity becomes scattered across parts of your brain and parts of the computer. (I don't know whether this is possible, or even whether it makes any sense; but it should be possible if the original uploading story is possible.) The procedure is then repeated with other parts of your brain one by one until all your mental activity (or all the mental activity that used to be yours) is going on in the computer and none is going on in your brain.

If the central dogma is true, it would presumably be possible to move a person from a human organism to a computer by means of such gradual uploading. If you could upload a person all at once, then you could upload a person gradually. But it doesn't look possible to construct a troubling duplication case involving gradual uploading–a case where there is no difference between moving you to a computer and merely creating a psychological replica of you there. And it would be quite a lot more difficult to construct a branching case, where there are two people, either of whom the friends of uploading would say was the original person were it not for the existence of the other.

Not that transhumanists will see this as good news. I doubt whether anyone thinks that gradual uploading is metaphysically possible but all-at-once uploading is not. There would have to be an explanation for this fact, beyond merely saying that all-at-once but not gradual uploading is subject to branching and duplication objections. It's hard to see what the explanation could be. In any event, it's clear that the metaphysical problems for the central dogma go deeper than the branching and duplication problems.

6. Material Continuity

I have said that the branching and duplication problems are symptoms of a deeper problem. What might this deeper problem be? If uploading really is metaphysically impossible, *why* is it impossible?

I think the answer is that you and I are material things: objects made up entirely of matter. That's certainly how it appears. That's why we're able to see and touch ourselves and other people. If we were immaterial, we should be invisible and intangible, which is very much *not* how it appears.

So we are material things. And a material thing cannot continue existing without some sort of material continuity. It must always be made up of some of the same matter–composed of some of the same material parts–that made it up at earlier times. A material thing can change all its parts: it can be made up of entirely different matter at different times. Owing to metabolic turnover, few atoms remain parts of a human being for long. But it cannot change all its parts at once. It cannot survive complete material *dis*continuity. It follows that you cannot move a material thing from one place to another merely by transferring information. You can't send a stone, or a shoe, or a dog as a message by telegraph (despite the joke in *Alice in Wonderland*). To move a material thing, you have to move matter–specifically, some of the matter making up that thing.[3]

But there is no material continuity in uploading. The person in the computer has none of the material parts of the human person. (Not in the usual "all-at-once" uploading, anyway.) The central dogma of transhumanism implies that you could send a person by telegraph–or, for that matter, written down in a letter. If I am right in saying that material things require material continuity to persist, then the central dogma is incompatible with our being material things.

We can make this more vivid by thinking about what sort of material things we might be. We appear to be animals: biological organisms. If you examine yourself in a mirror, you see an animal. The animal appears to be the same size as you–no bigger and no smaller. Like animals, we seem to extend just as far as the surface of our skin. Each of us seems to have the physical and biological properties of an animal: its mass, temperature,

[3] I believe that the material-continuity requirement derives from the further principle that material things must persist by virtue of "immanent" causation (Olson 2010). That is, they have to cause themselves to continue existing. Sometimes they need outside help–food, oxygen, medical care, that sort of thing–but the outside help can't do all the work. Corabi and Schneider (2012) argue that we cannot be uploaded because this would involve a gap in our existence. They say that material things cannot have such gaps, though I am unable to understand their argument for this claim. I suspect that if it is impossible, it's because it is ruled out by the immanent-causation requirement.

chemistry, anatomy, and so on. Nor is there any difference in behavior between a human animal and a human person. The appearance is that we *are* the animals in the mirror.

Our being animals is clearly incompatible with the central dogma. You cannot move a biological organism from a human body to a computer by scanning its brain and "uploading" the information thereby gathered. Scanning may leave the organism unharmed. Or it may damage it, perhaps even 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. It may remain unchanged, or be damaged or killed or completely destroyed, but it is not converted into information and transferred to the computer. You couldn't point to an electronic computer and say, "That thing was once a microscopic embryo composed of a few dozen cells."

So if you and I are organisms, it would be metaphysically impossible to upload us into a computer. Of course, we might be material things other than organisms. A few philosophers say that we are brains, or parts of brains. (Parfit 2012; see also Olson 2007: 76-98) Each of us is literally made up entirely of soft, yellowish-pink tissue and located within the skull. But it is no more possible to upload a brain into a computer than an animal. The scanning does not remove the brain from the head and convert it into information. The brain is a physical object, like a heart or a kidney. It may remain unchanged in the scanning process, or it may be damaged, or even completely destroyed by having its atoms dispersed, but it is not converted into information and transferred to the computer. You couldn't point to an electronic computer and say; "That thing was once a three-pound mass of soft tissue."

If the central dogma is true, then, it follows that we can be neither organisms nor brains. Not only could we not be organisms or brains once we have been uploaded, but we could not be organisms or brains even now. And not only are we not organisms or brains *essentially*. We are not organisms or brains even accidentally or contingently. The central dogma implies that a human person has a property that no organism or brain has, namely being uploadable into a computer by a mere transfer of information.

Suppose this attractive account of our metaphysical nature were true: we are biological organisms, or perhaps brains. What's more, all conscious beings, at the present time anyway, are organisms or brains. That would explain why a human person cannot be uploaded into a computer, contrary to the personal-identity assumption: because we are organisms or brains, and it is metaphysically impossible to move any material thing to a computer simply by transferring information.

7. The Pattern View

I have argued that we are material things, and that material things cannot persist without material continuity. As there is no material continuity in uploading, that explains why uploading is metaphysically impossible. (The same goes for *Star Trek* teleportation and Shoemaker's brain-state transfer.) There are two ways of defending the central dogma against this argument: to deny that we are material things, or to deny that material things require material continuity to persist. (I don't suppose anyone will argue that uploading is possible only in gradual cases where there *is* material continuity.) I will consider these proposals in turn.

To deny that we are material things is to deny that we are made up entirely of matter. And that is to say that we are partly or wholly made up of something else: we are (at least partly) *im*material things. And this is something that transhumanists often do say. Specifically, they often say that a person is a sort of *pattern*. Bostrom claims that in the future it will be possible for us to "live as information patterns on vast super-fast computer networks" (2016). Ray Kurzweil says that owing to the fact that 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. (Kurzweil 2006: 383)

And Daniel Dennett suggests that "what you are is that organization of information that has structured your body's control system." (1991: 430; see also 1978) Perhaps the very same pattern or form of organization could be instantiated or realized first in a biological organism and then in an electronic computer. And if this is possible, the scanning-and-uploading process that transhumanists imagine would be the way to do it. This may not be true of all patterns instantiated in the brain: those involving fluid dynamics or ion transfer across membranes are probably not transferable to an electronic substrate. But perhaps those patterns relevant to psychology are.

The proposal has to be that a person–the author of this paper, for instance–is literally a pattern of some sort. It is not merely that to be a person is to instantiate a certain sort of pattern, or that for a person to exist is for such pattern to be instantiated. These may or may not be sensible claims, but they do nothing to explain how a person could move from a human organism to an electronic computer. (They are compatible with our being organisms.) The suggestion is that we are not things that instantiate certain patterns, but that we are those patterns ourselves.

Could a conscious, thinking being be a pattern? The question is hard to think about because the word "pattern" is so nebulous. (I suspect that this lack of clarity is what has encouraged transhumanists to speak casually of our being patterns.) What sort of thing is a pattern? The assumption has to be that it is not a material thing of any sort, but rather something that can move from one material thing to another by a transfer of information. But that doesn't tell us much.

As far as I can see, a pattern would be a sort of property or relation: a universal. Different concrete objects, or collections of objects, can exemplify the same pattern, just as different flowers can have the same colour. All the copies of *Moby-Dick* in the original English have the same pattern of words and letters. (Or nearly the same. Let us ignore irregularities in typesetting, different locations of line and page breaks in different editions, and the like.)

The view that we are patterns, so construed, would solve the branching and duplication problems. If you were uploaded twice over, the result would be not two people, but only one: the same pattern would be present in two different computers. (Olson 2007: 146f.) There would be two *instances* of the pattern–that is, two physical things patterned in the same way–but there would be only one pattern in both. They would be the same person in the way that two physical volumes might be the same book–*Moby-Dick*, say. So double uploading would not have the impossible result that one thing is numerically identical with two things.

The proposal would solve the duplication problem by implying that a copy or replica of a person, if it instantiates the relevant pattern, *is* that person and not a replica. Both the Wittgenstein created by the British and the Wittgenstein created by the Austrians would be the original Wittgenstein born in 1889. Or more precisely, both physical objects would be instances of the same person: the solution assumes that neither physical object would itself be a person.

But the pattern view is impossible to take seriously. Suppose we ask which pattern a given person might be. If there are such things as patterns, this human organism now instantiates many of them. There is, for instance, the pattern consisting of the current orientation of my limbs, and the pattern formed by the flow of material through my gut. Which pattern am I? Since I am conscious and thinking, I must be the one that instantiates those mental properties. The pattern view presupposes that of all the patterns instantiated here, one of them, and only one, can think. That's because there is just one thinking being here, namely me. But which of those patterns is the one that thinks? Of all the patterns the organism instantiates, what could make just one of them conscious? I have no idea how

to answer this question. It's no good saying that to be conscious or intelligent *is* to instantiate a certain pattern. Although that may be true, it would imply that the organism was conscious, since it is the thing instantiating the pattern. That would make typical human organisms conscious and intelligent, yet not uploadable–precisely what the pattern view was meant to avoid. The proposal has to be that no material thing could possibly have any mental property.

But perhaps the most obvious problem for the pattern view is that universals don't *do* anything. They don't change. And this prevents them from thinking or being conscious. When we speak of changing the pattern or arrangement of chairs in the room–from square to circular, say–we mean rearranging the chairs so that they instantiate a *different* pattern from their current one. A single pattern cannot be first square and then circular. It can change only in the way that the number seventeen changes by ceasing to be the number of chairs in the room when we move one next door: mere "Cambridge change," as they say. A universal cannot undergo any real, intrinsic change.

But if I know anything, I know that *I* undergo real change. I am sometimes awake, for instance, and sometimes asleep. That I change intrinsically follows from the fact that I am conscious and thinking. No person–even a computer person–could be a pattern. A thing that changes can at best be a particular *instance* of a pattern and not the pattern itself: a concrete thing that is patterned or organized or arranged in that way. The claim that a person is an instance of a pattern is entirely harmless. Every concrete object is an instance of some pattern or other (still supposing that there are such things as patterns). But again, the claim that we are instances of patterns tells us nothing about how we could be uploaded into a computer.[4]

8. The Constitution View

Turn now to the proposal that we can survive complete material discontinuity despite being entirely material things. One view of this sort incorporates the thought that a human person is not an organism, but rather a material thing "constituted by" an organism. Each of us stands to an organism in the way that a clay statue stands to the lump of clay making it up. A human person is made of the same matter as the organism we might call its body, and physically indistinguishable from it. But the person differs from the organism in its modal properties: the person, but not the organism, persists by virtue of psychological continuity. So in Shoemaker's brain-state transfer story, a person would be constituted first by one organ-

[4] For more on the pattern view, see Olson 2007: 145-149.

ism and then by another, much as a statue that got an arm replaced would be constituted first by one lump of clay and then by another. And perhaps in uploading, a person could cease to be constituted by any organism, and come instead to be constituted by some part of a computer.[5] This need not imply that *all* material things can survive without material continuity. It might be impossible for an organism or a lump of clay. But material things of our sort can.

There is a large and ongoing debate over the merits of the constitution view, independent of whether it would allow uploading.[6] But the view is unlikely to appeal to transhumanists. For one thing, it does nothing to explain *how* it is possible for a material thing to survive without material continuity. If it seems absurd to suppose that a thing made entirely of matter could be sent as a message by telegraph or dictated over the phone, the proposal tells us nothing about why this appearance is misleading. It says, of course, that personal identity over time consists in some sort of psychological continuity, generously construed so that it does not require material continuity. But this simply asserts that material continuity is unnecessary, and does nothing to address the strong conviction to the contrary. What's more, the claim is entirely independent of the constitution view. If it's a sensible thing to say, it's sensible whether or not we are constituted by organisms.

Nor does the proposal suggest any solution to the branching and duplication problems. If uploading could bring it about that I ceased to be constituted by an animal and became constituted instead by a computer, then it could apparently bring it about that I became constituted simultaneously by one computer and also by another, making me numerically distinct from myself. And there would appear to be no difference between a computer's constituting me as a result of uploading, on the one hand, and a computer's constituting someone else just like me, on the other, and thus no difference between a person and a mere copy of that person.

9. The Temporal-Parts View

The best way of defending the central dogma may be to appeal to the ontology of temporal parts.[7] It consists of two principles. First, all persisting things are composed of arbitrary temporal parts. A temporal part of some-

[5] Both Baker and Shoemaker believe that we are constituted by organisms and that we can survive without material continuity (Baker 2005; Shoemaker 1984: 108-114, 1999). Given the AI assumption (which Baker accepts; cf. 2000: 109), it follows that I could become constituted by a computer through uploading.

[6] For a summary, with references, see Olson 2007: 48-75.

[7] This is a difficult topic. I discuss it at greater length in Olson 2007: 99-128.

thing is a part of it that takes up "all of that thing" at every time when the part exists. Barry Manilow's nose is a part of him, but not a temporal part, because it doesn't take up all of him while it exists. His adolescence or his first half, though, if there are such things, would be temporal parts of him. A temporal part of something is exactly like that thing at all times when the part exists. It differs from the whole only by having a shorter temporal extent. To say that persisting things are composed of *arbitrary* temporal parts is to say that for any period of time when a thing exists, there is a temporal part of it existing only then.

The second principle is unrestricted composition: for any entities whatever, there is a larger thing composed of them. (Some things, the *x*s, compose something $y =_{df}$ each of the *x*s is a part of *y*, no two of the *x*s share a part, and every part of *y* shares a part with one or more of the *x*s.) So if there are such things as Barry Manilow's nose, Plato's fourth year, and Yugoslavia, then there is also an object scattered across space and time that is made up of those three things. Both principles are, of course, highly controversial. Together they imply that every matter-filled region of spacetime is exactly occupied by a material thing. This is what Quine meant when he said that a physical object "comprises simply the content, however heterogeneous, of some portion of space-time, however disconnected and gerrymandered." (1960: 171)

It follows from the principle of arbitrary temporal parts that I have a temporal part extending from the beginning of my existence until midnight tonight, and that my computer has a temporal part extending from that time until the computer's demise. And it follows from unrestricted composition that there is something composed of these two objects: a material thing, given that both I and my computer are material things. It is conscious and intelligent until midnight tonight, when it "jumps" discontinuously from me to the computer. From then on it is not conscious or intelligent. (Splendid though my computer is, its powers are limited.) If my computer really did have the right mental capacities, though, then the being jumping from me to it would remain conscious and intelligent. In fact such a being would make this jump at every moment at which both the computer and I are conscious, with or without any sort of "uploading"–that is, any transfer of information from the organism to the computer. That's because the computer and I are each composed of arbitrary temporal parts, and any two of them compose something. Any pair consisting of one of my temporal parts and one of my computer's, provided they don't exist simultaneously, will jump from one of us to the other.

So according to the ontology of temporal parts, it is perfectly possible for a material thing–even a conscious, intelligent one–to persist without

material continuity. It does not follow from this, however, that a *person* could move from a human body to a computer. To secure this claim–the personal-identity assumption–such beings would have to count as people. And on the temporal-parts ontology, having the mental capacities characteristic of personhood–intelligence, self-consciousness, and the like–does not suffice for being a person. Many of my temporal parts, such as the one that extends from midnight last night till midnight tonight, have those mental capacities but are not people. No person now writing these words is going to perish at the stroke of midnight, without any injury or other disruption of his mental or physical activities. At any rate, few temporal-parts theorists think so. (Sider [1996] is an exception.) Not just any rational and self-conscious being is a person.

We can see this point by noting that the ontology of temporal parts entails the existence of a thing composed of the temporal part of me extending from my beginning till midnight tonight and the temporal part of you extending from that time till your demise: a conscious, intelligent being jumping from me to you. But this being is not a person, and its existence is of no practical or metaphysical interest. If I knew that I was going to be shot at dawn, the conviction that this being was going to survive that event would be no more comfort me than the thought that you were going to survive it.

So the temporal-parts ontology implies that conscious, intelligent beings could move from human bodies to computers by uploading. There is no metaphysical mystery about this–or at least none beyond that inherent in the temporal-parts ontology itself and the AI assumption.

The proposal would also solve the branching and duplication problems. Suppose my brain is scanned (and thereby erased) and the information gathered is uploaded simultaneously into two computers. Two people emerge from the process. Both, temporal-parts theorists can say, would be me. How could two things be one thing? The reply is that in this case there are two people all along, who share their pre-upload stages but not their post-upload stages. (Call the short-lived temporal parts of people "person stages.") These people begin to exist when I do and share all the events of my life until the uploading takes place. During that period there is no difference between them. But afterwards they live in different computers and lead independent lives. This is a consequence of the claim that there is a being composed of my pre-upload stages and the post-upload stages of the one computer, and also a being composed of my pre-upload stages and the post-upload stages of the other computer, together with the assumption that such stages are connected in the way that makes for personal identity over time–that is, that makes them compose a person. The two people are

like railway lines that share their tracks for part of their length and diverge elsewhere.

What about the duplication problem? What would be the difference between bringing Wittgenstein himself back to life, by programming a computer with the psychological information from his brain, and creating a psychological replica of him by that means? According to the temporal-parts ontology, there is no deep metaphysical difference between originals and replicas. Suppose we somehow produced a computer person psychologically identical to Wittgenstein as he was shortly before his death. The temporal-parts ontology would imply that there are two conscious, intelligent beings in the computer, insofar as the intelligent computer stages are parts of two such beings. One was born in 1889 and wrote the *Tractatus Logico-Philosophicus*. The other began to exist only just now. They share their current stage, but the 1889-to-1951 stages are parts of the first and not of the second. If the first being counts as a person, then we have resurrected Wittgenstein himself. If the second is a person, then we have merely created a replica of him. (Requiring a person to be a *maximal* aggregate of appropriately interconnected stages—a thing, each of whose stages is appropriately connected to every other, but which is not a part of any larger such thing—would rule out their both being people.) But which of these is the case is not a metaphysical question, but simply a matter of how we use the term "person."

According to the temporal-parts ontology, then, conscious, intelligent beings could move from human bodies to computers via uploading. And these beings would be people, vindicating the personal-identity assumption, just if the stages of those beings would relate in the way that would amount to their composing a person. What relation is this? Transhumanists will say that it is some sort of psychological continuity or connectedness. Perhaps a person is a maximal aggregate of psychologically interconnected person stages: that is, a being composed entirely of person stages, each of whose stages is psychologically connected to every other, and which is not a part of any larger such being. (Lewis 1976) And we might say that two person stages are psychologically connected just if the mental properties of one of them depend causally in the right way on those of the other.

It's clear that the post-upload stages of a computer person could have mental properties that depend causally on those of the pre-upload stages of human people. But would they depend in the right way–the one that would make the beings who move from human being to computer count as people? That looks doubtful. An attractive thought is that stages are parts of the same person only if they are connected by relations of practical concern: if one has "what matters" to the other. (Parfit [1984: 262] calls this

connection "relation *R*".) In other words, a future person is me only if I now have a reason to care about his or her welfare then–a reason I should have even if I were completely selfish and would not lift a finger to save my own mother from unbearable pain. This may also imply that that future person would be morally responsible, then, for the things I do now (in the absence of the usual excuses, such as insanity), that he or she would then deserve compensation for my efforts now, and so on.

Transhumanists are likely to accept this requirement. Their claim that we could become computer people is not meant to be of merely theoretical interest. They think it would matter to us, practically speaking, if we were uploaded into computers: it could benefit us, and we have a reason to try to bring it about. If we could become computer people but their welfare would be of no practical importance to us, then we should have no selfish reason to upload ourselves, no matter how wonderful the life of a computer person would be. Uploading ourselves would be no better, for us, than creating psychological duplicates of ourselves in computers.

In fact, transhumanists would rather say that computer people could have what matters practically to us without being us than that computer people could be us without having what matters. That is, if the personal-identity assumption turned out to be false and it was metaphysically impossible for us to become computer people, they would retreat to the claim that computer people could at least have what matters: they could bear to us those relations of practical concern that give us a reason to care about our own future welfare. Even if we cannot literally be uploaded, they will say, it could be that as far as everything we care about is concerned, it's as good as if we could. My worries about the personal-identity assumption are of interest only to metaphysicians. The rest of us can ignore them.

So according to the temporal-parts ontology, we could be uploaded into computers only if beings that move from human bodies to computers via uploading could count as people. And they could count as people only if each of their stages has the mattering relation to every other, or at any rate only if their post-upload stages have what matters practically to their pre-upload stages. Is this the case? It doesn't seem so. Consider once again the case of nondestructive uploading. Suppose I am kidnapped by bad people, who are going to scan my brain and upload the information, resulting in both a computer person psychologically just like me and a human person entirely like me (and materially continuous with me to boot). Then they are going to torture one of these people. The magnitude of the suffering will be the same in either case. But for some reason they allow me to choose which person is tortured: the human person with my body or the computer person. (Suppose I accept the AI assumption: I don't doubt that they could create and torture a computer person.)

If uploading preserves what matters, I ought to be indifferent. But I should be anything but indifferent. Even if I were completely selfish, I would far rather that the computer person be tortured than the human person. I suspect, in fact, that if I were completely selfish I should be indifferent about the welfare of the computer person. My only concern would be the welfare of the human person. And I doubt whether these attitudes are peculiar to me.

Or imagine that the bad people work out how to scan people's brains remotely without their noticing. They then upload the information from the scan into a computer, creating someone psychologically identical to the unsuspecting victim as she was when she was scanned. This being is then tortured. Suppose the bad people have been active in my neighbourhood, and I believe there is a real chance that they will scan my brain tonight and torture the resulting person. If uploading preserves what matters, I ought to be just as worried about this as I should be if I thought there was a real chance that the human person who will wake up in my bed tomorrow will be tortured. But I should find the second case far more worrying.

Someone might suggest that *destructive* uploading preserves what matters practically, even though nondestructive uploading does not. A computer person produced by scanning and uploading the information in my brain would have what matters to me if, but only if, there is not also a human person then who is both psychologically and materially continuous with me. And this is not because the computer person *is* me just if no such human person issues from the procedure: we are assuming that a person moves from my body to a computer in either case. The suggestion is that the computer person would be me (by sharing my current stage), but whether I should have any selfish reason to care about his welfare depends on what *other* people existing after the procedure would also be me (by sharing my current stage). But no philosopher I know of has ever held this view. Parfit's famous arguments in Part 3 of *Reasons and Persons* presuppose that what matters cannot depend on what we might call extrinsic factors, and none of his many critics have questioned this assumption.

It appears, then, that uploading does not preserve what matters practically. Assuming that stages are parts of the same person only if they are connected by relations of practical concern, it does not look as if a person could move from a human organism to a computer by uploading. The procedure may move *some* material thing from a human being to a computer–this is guaranteed by the ontology of temporal parts–but not a person. The personal-identity assumption looks false even given the temporal-parts ontology.

So it looks rather doubtful whether the temporal-parts ontology can save the central dogma of transhumanism. For the same reason, it looks doubtful whether computer people could have what matters to us in identity: whether having psychological duplicates in computers would be just as good for us, practically speaking, as literally moving there ourselves. But I'm not very confident about this. It may be that I am simply wrong about what matters practically–about what would be in my own interest–and that my reactions to the imagined cases are mistaken. In any event, transhumanists have work to do.

REFERENCES

Baker, L. R. (2000). *Persons and Bodies: A Constitution View*. Cambridge University Press.

Baker, L. R. (2005). "Death and the afterlife." In W. Wainwright (ed.) *Oxford Handbook for the Philosophy of Religion*. Oxford University Press: 366-39.

Bostrom, N. (2001). "What is transhumanism?" <http://www.nickbostrom.com/old/transhumanism.html> accessed 9.8.16.

Bostrom, N. *et al.* (2016). Transhumanist FAQ, 3.0. <http://humanityplus.org/philosophy/transhumanist-faq/>, accessed 9.8.16.

Chalmers, D. (2010). "The singularity: A philosophical analysis." *Journal of Consciousness Studies* 17 (9-10): 7-65.

Corabi, J. and Schneider, S. (2012). "The metaphysics of uploading." *Journal of Consciousness Studies* 19 (7-8): 26-ff.

Dennett, D. (1978). "Where am I?" *In Brainstorms*. MIT Press.

Dennett, D. (1991). *Consciousness Explained*. Little, Brown.

Kurzweil, R. (2006). *The Singularity is Near: When Humans Transcend Biology*. Duckworth.

Lewis, D. (1976). *Survival and identity*. In A. Rorty (ed.) *The Identities of Persons*. University of California Press. (Reprinted in his *Philosophical Papers* Vol. I, Oxford University Press, 1983.)

Olson, E. (2007). *What Are We?* Oxford University Press.

Olson, E. (2010). "Immanent causation and life after death." In G. Gasser, (ed.) *Personal Identity and Resurrection*. Ashgate: 51-66.

Olson, E. (2015). "Life after death and the devastation of the grave." In M. Martin and K. Augustine, (eds.) *The Myth of an Afterlife*. Rowman & Littlefield.

Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

Parfit, D. (2012). "We are not human beings." *Philosophy* 87: 5–28.

Quine, W. V. O. (1960). *Word and Object*. MIT Press.

Shoemaker, S. (1999). "Self, body, and coincidence." *Proceedings of the Aristotelian Society, Supplementary Volume* 73: 287-306.

Shoemaker, S. (1984). *Personal identity: A materialist's account*. In S. Shoemaker and R. Swinburne *Personal Identity*. Blackwell.

Sider, T. (1996). "All the world's a stage." *Australasian Journal of Philosophy* 74: 433-53.

van Inwagen, P. (1978). "The possibility of resurrection." *International Journal for Philosophy of Religion* 9: 114-121.