



Personal Identity and Uploading

Mark Walker

Richard L. Hedden Chair of Advanced Philosophical Studies

Department of Philosophy

New Mexico State University

mwalker@nmsu.edu

Journal of Evolution and Technology - Vol. 22 Issue 1 – November 2011 - pgs 37-52

Abstract

Objections to uploading may be parsed into substrate issues, dealing with the computer platform of upload and personal identity. This paper argues that the personal identity issues of uploading are no more or less challenging than those of bodily transfer often discussed in the philosophical literature. It is argued that what is important in personal identity involves both token and type identity. While uploading does not preserve token identity, it does save type identity; and even qua token, one may have good reason to think that the preservation of the type is worth the cost.

1. Uploading: prospects and perils

You arrive at one of the thousands of kiosks run by the late twenty-first century's largest corporation: U-Upload. With some trepidation you step into the superscanner. There is a slight hum as it inventories the molecular building blocks of your brain. Your brain is destroyed in the process, but you are not dead – or so the marketing materials from U-Upload claim. For information about the building blocks, along with a general program that describes the fundamental laws of molecular interaction, is uploaded to the shiny new robotic brain you purchased (Sandberg and Boström 2008). For your friends and family, a few terrifying moments pass before the robotic body stirs. To their relief, your first words are: "It's me. I made it." You then go on to crack a joke – just as your family and friends have come to expect of you. Of course you have changed in some respects: gone is your human carbon-based body. Now you experience the world through camera eyes and microphone ears, you dance the fandango with robotic legs and speak

through a voice synthesizer. But it is still you. You have migrated to a silicon substrate: you have been uploaded.

At least that is one interpretation of these events. The contrary construal is that although a robot was created that acts and talks like you used to, claims to have your memories, and indeed, claims to be you, this robot is not you. You are dead. You died when your brain was destroyed during the scanning process.

If, like me, you think that uploading is possible (at least in principle), and so you hold that the first interpretation of these events is correct, then you must hold true the following three theses:

[1] Computers are capable of supporting the important properties constitutive of personal identity, e.g., thought and consciousness.

It is clear that uploading will not preserve all properties we associate with *Homo sapiens*, e.g., basic facts about the human digestive system are not likely to be preserved in uploading to a robotic body. But these facts are not typically thought to be important for personal identity. Candidates for important properties include thought, consciousness, emotions, creativity, aesthetic experience, sensory experience, empathy and so on. For the most part, the question of which properties are important is not as serious as it may first seem, since uploading promises to preserve the essential aspects of the brain and nervous system, which overlap with the usual lists of important properties for identity.

A famous challenge to thesis [1] is made in Searle's Chinese Room argument (Searle 1980). It is beyond the scope of this paper to explore this argument; suffice it to say that if Searle is correct, then [1] may be false. For Searle thinks that a computer can never consciously think merely in virtue of instantiating a computer program, and the uploading process seems to be one of merely instantiating a computer program (Agar 2010, 2011).

[2] It is possible to capture the information necessary to emulate the important properties of individual humans.

The technical challenge of thesis [2] is to capture the information in all parts of the brain in a manner that preserves the relevant information. Clearly this won't be easy. If we slice off layers of your neurons, and record the information of each layer, the lower layers will change (due to trauma or death).¹ If we flash freeze your brain, we may destroy some essential information. Philosophical questions arise as to whether the information encoded in the brain is sufficient to account for all the relevant properties. For example, consider a dualist who believes that we have souls in addition to brains, and much of what is morally important (e.g., conscious thought) resides in the soul. If the dualist is right, then scanning your brain could never be sufficient, for it would be necessary to scan your soul to access at least some of what is important. If it is unlikely that we will be able to scan souls, there will be an insurmountable obstacle to uploading. Notice how theses [1] and [2] may differ on this point: a dualist could consistently hold that a computer might have a soul; it is just that if computers have souls, it is not because we obtained the soul-building information from humans. (Perhaps God implants souls in humans and computers.)

[3] It is possible to survive the uploading process.

To see how [3] differs from [1] and [2], imagine that at some point in the future we have created computers of sufficient complexity that it is agreed that they have the same morally relevant properties as humans: these advanced computers think and are conscious, they are accorded rights, and the scanning problem has been solved so that we are able to scan the brain in such a way that we are not worried about

loss of information. None of this answers the question of whether you have been preserved during uploading or whether uploading merely makes a very good copy of you. The worry that only a copy is created is often fueled by this thought: the information about the building blocks of an individual human brain could be uploaded to multiple computers with robotic bodies. The number of copies of a person is limited only by the available computing power. If an individual can be uploaded once, then it seems the same individual could be uploaded twice into separate computers, and indeed, billions of the same individual all embodied in separate robotic bodies could be created.

This quick survey of the conceptual terrain suggests that there are substantial philosophical (not to mention technical) obstacles to uploading. To make the discussion manageable, I will focus on thesis [3], and assume without argument that [1] and [2] have been resolved in favor of uploading. So our question is this: assuming that computers can be conscious, have memories, and (robotic) bodies, and assuming that it is possible to scan and capture all the information of a human brain, does uploading preserve personal identity?

I will argue that uploading does preserve personal identity; at least identity of a certain sort.

2. The equivalency thesis

The fact that we are assuming that computers are capable of embodying all the same type of properties necessary for personal identity means that we can make use of the equivalency thesis:

Equivalency thesis: If it is possible for an individual to survive migration from a carbon to a carbon body, then it is possible for individuals to survive migration from a carbon to a silicon body.

To spell this out, I'll say first what I mean by occupying different human bodies, and then say what use the equivalency thesis will serve for us.

Let us start by considering a familiar fictional example of people switching human bodies, i.e., carbon-to-carbon transfers. One of my personal favorites is a schlocky episode in the original *Star Trek* series. Captain Kirk finds himself in the body of his jilted ex-lover, Dr. Janice Lester, after an alien "personality swapping" device is used on him. She, jealous of his power, takes control of his body, and, what is worse, his spaceship. This plot device has been used numerous times since, including in the movie *Freaky Friday* where a mother finds herself in her teenage daughter's body and vice versa. These works of fiction are premised on the idea that whatever makes individuals the individuals they are is only contingently related to the bodies that they find themselves in. Captain Kirk grew up in a male body, but we are asked to believe that, at least for a short while, he inhabited a female body.

In Kirk and Lester's body swap, the idea is helped along by the visual effects (such as they were in the 1960s) that showed what apparently we are to understand as soul swapping. (A soul, it turns out, looks much like a translucent version of one's body. Who knew?) We do not need to have recourse to the idea of souls, however. Imagine the scanner used to encode all the relevant biochemical information from a brain was used to scan both Kirk's and Janice's brains. Nanobots – nanoscale robots – then rearrange the biochemicals in each brain to encode the relevant memories, personality and intellectual abilities and so on. This differs then from brain swapping, because each brain is reorganized using nothing but the locally available biochemicals. Here the information is uploaded to a different human body rather than a computer. Using this procedure, it makes perfect sense why Captain Kirk's body would act much like we would expect Dr. Janice Lester to act, and vice versa.

There are a couple of reasons for invoking the equivalency thesis. The first is so that we are not misled by a new form of racism: substratism (Walker 2006). Substratism is the view that one's substrate is inherently superior to that of other substrates along the lines that racists think their race is inherently superior to some other race. In the present case, it would suggest the idea that carbon-based humans are inherently more morally worthy than silicon based beings. Consider the fact that we would not accept this argument: it is not possible for persons to migrate from one body to another because then it would be possible for people of skin color X to move to bodies of skin color Y, and Y skin color is morally inferior. We want to avoid the same bad argument in considering moving from one substrate to another. Notice that this does not beg the issue at hand, since it is possible to say that having a certain substrate (or even skin color) is constitutive of my identity; it merely prohibits saying that this property in itself makes for moral superiority.

The second is that it makes directly relevant an enormous amount of philosophical effort that has gone into exploring the possibility of carbon-to-carbon transfers. The question of carbon-to-silicon transfers thus may piggyback on this effort.

3. Personal identity: psychological and somatic accounts

Historically, there are two main schools of thought about what is required for personal survival; the psychological and somatic approaches (Olson 2002). Derek Parfit's famous thought experiment may serve as illustration:

I enter the Teletransporter. I have been to Mars before, but only by the old method, a space-ship journey taking several weeks. This machine will send me at the speed of light. I merely have to press the green button. Like others, I am nervous. Will it work? I remind myself what I have been told to expect. When I press the button, I shall lose consciousness, and then wake up at what seems a moment later. In fact I shall have been unconscious for about an hour. The Scanner here on Earth will destroy my brain and body, while recording the exact states of all of my cells. It will then transmit this information by radio. Travelling at the speed of light, the message will take three minutes to reach the Replicator on Mars. This will then create, out of new matter, a brain and body exactly like mine. It will be in this body that I shall wake up. (Parfit 1987, 199)

Those that hold the psychological account of personal identity will tend to endorse the view that one survives teletransportation. For the psychological account says that what is essential for survival is continuity of psychological states such as memory, beliefs, desires and personality. John Locke, an early proponent of this view, famously described personal identity in terms of psychological continuity, within an analysis of personhood as consisting in existence as "a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places..." (Locke 1975). The person on Mars who awakens will claim to remember being Derek Parfit, and to have memories and a personality that are psychologically indistinguishable from the person on earth whose body was destroyed. Locke, then, would say that Parfit survived teletransportation.

Somaticist accounts suggest the survival of a particular body is critical for personal identity over time. Since the body on Earth is destroyed during the scanning process, Parfit ceases to be. A different person will awake on Mars. This person will of course have psychologically indistinguishable memories and personality to those of the late Parfit, but this person will not be Parfit. The new person will be but an infant in terms of chronological age: only a few minutes old.

We will think of “somaticism” as the view that continuity of one’s body is necessary for personal identity from one time to the next.² There are two ways that one might be a somaticist: one can believe that bodily continuity is necessary but not sufficient, or that it is necessary and sufficient. One easy case to distinguish these two is as follows: a piano falls on your head, and causes you to go into a permanent vegetative state. Your relatives discuss whether to “pull the plug.” Those who think that bodily continuity is necessary but not sufficient may say that you no longer exist, but your body continues to exist. Those who think that bodily continuity is necessary and sufficient will say that you continue to exist, albeit your cognitive capacities are non-existent. Both views qualify as “somaticism” in our sense.

In my extremely limited and informal survey of students and friends, most would be unwilling to step into Parfit’s Teletransporter.³ The usual response is that it is equivalent to committing suicide: the person here is killed, and a new duplicate is created. Nothing of the original survives. Locke and others who endorse the psychological account would retort that this is just an irrational attachment to a certain set of molecules. If a molecule-for-molecule identical copy is made, then it seems irrational to prefer one set of molecules to another. Parfit compares this to an attachment to a wedding ring: there may be sentimental value in having the original rather than a molecule-for-molecule identical copy, but such attachments are “merely sentimental” and have nothing to do with personal identity.⁴

The reluctance to use the transportation device seems hard to explain other than by the fact that people hold, at least implicitly, to somaticism. This provides a robust challenge to psychological accounts in general, and a challenge to uploading in particular. In the next three sections I will offer arguments against somaticism.

4. Against somaticism: the big stroke

The Vorlons,⁵ a mysterious and intellectually advanced alien species, make this offer: you can have an original undiscovered play by Shakespeare written in his hand, or a copy of the play made by one of his lackeys. You salivate at the joy this will bring to the world (not to mention the fame and fortune it will bring you personally). Since you can have only one, the choice, it seems, is a no-brainer. You should opt for the one written by the bard’s hand. But now consider this variant: the Vorlons tell you that the text written in Shakespeare’s hand is missing the last two pages, while they assure you the copy written by the lackey is a perfectly faithful reproduction of all the words in the original. While it would be great to have both, you reason that the most important thing is the play itself be preserved, not Shakespeare’s handwriting. The copy here is in some sense better than the original because the original has been damaged. This tips the scales in favor of the copy, because while being written by the bard’s own hand is good, having the whole play is even better.

We can apply this lesson to thinking about personal identity. The Vorlons, with their ability to see into the future, say the news is grim. In less than twelve hours you will have a massive stroke that will cause you to lose many of your memories and some mobility, and impair your intelligence. Your stroke will not be as bad as some: the damage from the stroke will not leave you completely cognitively impaired, but you will no longer be able to work as an academic. You will have to find some relatively mindless job befitting your new level of intelligence, perhaps in academic administration. Friends and family will say that your once keen memory has been dulled such that your memory is now fuzzy, and you seem to remember the most superficial things. It is a shame, and totally unexpected at your young age. Even with their immense power, there is nothing the Vorlons can do to prevent the stroke. They provide a radical alternative: creating a perfect replica of you – down to the molecular level – with the exception that the problems with the arteries to your brain will be fixed in the body replica. They insist, however, that only one body can survive. You must choose tonight whether the replica or your current body survives.

Most people I've asked about this would rather see the replica survive, for the replica best embodies what is most important about you: your memories, your personality, your beliefs and desires. None of this is to say that the loss of one's body is trivial. One can be quite attached to one's body; but, when given this tragic choice, more of what is essential to you as a person survives in the replica.

Obviously, this example is structurally similar to Parfit's, but with one big exception: what is gained by having the replica survive is much more significant in this case than in the Teletransporter to Mars case. Parfit offers the incentive of avoiding three weeks of space travel. (We might not even sacrifice our original wedding ring for an exact replica if the benefit is merely avoiding three weeks in a spaceship). Here the incentive is the possibility of not having one's life radically altered by the stroke. The attachment to one's body does not seem worth the cost in this case.

Perhaps it might be remonstrated thus: "If I survived in a brain damaged state, I would be a terrible burden on my family and the world. It would be better for my family and the world that I died and a replica replaced me." To avoid this objection we can simply stipulate that the decision is to be entirely selfishly motivated, and that we know this about your preferences: you would rather survive a stroke than not survival at all. So, if the Vorlons did not offer you a chance to survive as a replica, you would rather live after the stroke than die. If the choice is still to have your present body die (the one with the bad arteries), then this can only be explained by thinking that you will survive as a replica.

It may be thought that even the most selfish person might prefer death if it meant something else he or she valued might result, e.g., you value the finishing of your novel more than you value your own life. If a replica of you can better realize this project, then it is consistent with selfishness to prefer death to oneself for the sake of the great unfinished novel. Again, we may simply stipulate around this objection. We may say simply that what you want most is for *you* to finish writing the novel, not someone else. If you die, you would rather it remain the "great unfinished novel" than be finished by someone else. If this is your most important desire, then preferring the stroke body's death cannot be explained away by the thought that what you wish for is the completion of your projects.

It is worth noting that not all somaticists are likely to be convinced by this example.⁶ But it should convince a few, and points out one of the heavy costs of somaticism.

5. Against somaticism: retrospective replicas

In this section we will examine an alternative explanation to somaticism for why people might be reluctant to use Parfit's Teletransporter, namely, fear of the unknown. Notice that Parfit's case is prospective: he asks us to imagine the decision to walk into the scanner with the hopes of being teletransported. The thought is that fear of the unknown may be muddying the waters here. That is, perhaps it is this fear of the unknown, rather than a commitment to somaticism, that explains the reluctance to use the Teletransporter. We can test this thought by considering a retrospective rather than a prospective version of a replication scenario.

Suppose that every night when people sleep their bodies (including their brains) are scanned by a swarm of nanobots and a molecule for molecule identical body is beamed from a hidden alien spaceship in orbit; the old body is vaporized in a manner that is undetectable by the human eye. Scientists discovered this fortuitously: physicists noticed a spike in neutrino levels every time psychologists in the adjoining lab conducted sleep experiments. Intrigued, scientists built a chamber to isolate subjects from neutrino

influences and then had test subjects sleep in the chamber. Once the experiment was initiated, a hologram of a Vorlon appeared in the lab and spoke thusly:

We are an ancient race known as the “Vorlons.” We battled another species, the “Shadows,” just as your species was beginning to evolve on this planet. One of the toxic effects of our war was a type of radiation that kills all higher intelligences within three days. We have no way of eliminating the radiation, but we have left advanced technology to recreate your bodies from different molecules every day so that the radiation will not harm you. We left the galaxy eons ago. You are hearing this message now because you have advanced technologically to the point where you can detect our technology. If you interfere with our replicator technology, you will quickly die of radiation poisoning.

What should we make of this? It is clear that dismantling it is out of the question since all humans will die within three days. If you are a somaticist, you must conclude that you have been alive only for a very short while. In fact, you have existed only since last night. After all, the physical continuity of one’s body has lasted only this length of time. However, most of us, I think, would conclude the opposite. That is, that we have existed for years: that we do not cease to exist every night and a new person comes into being.

This example may not be a decisive refutation of somaticism, but it does at least pull out one pillar of support. Somaticists ask why so many would be reluctant to step into the Teletransporter that Parfit describes, intimating that our reluctance has to do with the fact that our bodies will not survive. The retort, suggested by this example, is that the reluctance is explained more simply as a fear of the unknown. Contrariwise, the somaticist must now explain how so many people could be mistaken about their own identity retrospective case; after all, it seems very likely that, upon learning about the Vorlons’ technology, most would conduct their lives as if they hadn’t just come into existence that day. Who is going to say such things as: “I do not have to look after these children you call mine: how can I have children if I myself was born today. I can’t use this driver’s license, it is someone else’s – I was just born today. I’m not qualified to teach any classes: a postgraduate degree is required, which takes years to earn, and I was just born today?”

6. Against somaticism: practical ethics

Many of the problems of personal identity are simply extensions of the more general problem of identity; for example, a meteor falls on a family dining room table smashing exactly half of it beyond repair. The other side is virtually unscathed. Expert carpenters are brought in to fix the missing half. Has the table survived? Most say yes. Here is a variant on the story: siblings fight over who gets to inherit the family dining room table. In the end they saw it exactly in half and each sibling hires expert carpenters to replace the missing half. Did the original table cease to exist when it was cut in half by the siblings? If we say yes this seems to conflict with the original intuition that a table can survive the loss of half of its material. If we say no, then it seems we have the impossible situation where numerically distinct tables (each owned by one of the siblings) are not in fact numerically distinct. As intimated above, we can construct parallel cases for personal identity. My point here is that there are complex metaphysical issues of which at least some personal identity issues look to be merely specific instances of more general problems.

Since issues of personal identity are acknowledged by all to be deeply contested, and since they may be intimately intertwined with the more general and equally contested issue of identity, it looks like we won’t be able to resolve these issues anytime soon. Hence, we are left in a quandary about how to proceed. To emphasize, let us suppose that despite the fact that the preponderance of reasons seem to be

against somaticism, imagine that the metaphysical reasons for and against somaticism are exactly balanced.

Does this imply that we should be neutral on the issue? I think not. It may be that there is a further court of appeal to decide the issue, specifically, practical ethics. That is, the suggestion is that if our metaphysical arguments and intuitions cannot decide the metaphysical issue of personal identity, it is permissible to decide the issue on non-metaphysical grounds. I won't argue for this claim here, as it will take us too far into the meta-philosophical issue of how different areas of philosophy, in this case, metaphysics and practical ethics, are related. Suppose, for the moment, it is true that practical ethics can tell us something about metaphysical issues, it is then reasonable to ask: What does practical ethics tell us about the issue of personal identity?

Imagine two persons, McCoy and Hatfield, who want to kill one another. They are co-inventors of the first replicating machine. McCoy thinks it should be used on humans, Hatfield believes that it never should be so employed. McCoy believes in the psychological continuity thesis of personal identity, whereas Hatfield believes in somaticism. How should we reason about personal identity in terms of what is good for society? We can imagine two possibilities: society adopts for, social and legal purposes (its "public norm" for short), somaticism or psychological continuity. Which is better for society?

Consider first using somaticism as the public norm. McCoy could kill Hatfield and then hop in the replicating machine. We would be forced to say, because we have adopted somaticism as our public norm, that McCoy is dead and the replica of McCoy (call this person "McCoyson") is a different person. Since McCoyson was born after the crime, McCoyson cannot be responsible for the crime. (We have long abandoned the idea that one can inherit personal responsibility for the sins of one's ancestors). This crime would be ruled a murder-suicide in a somaticist jurisdiction. Of course McCoy then has every reason to commit the crime, as he does not believe in somaticism. We can easily imagine that the number of murder-suicides would greatly increase. Indeed, mass murder hardly seems out of the question. If you are looking for a job in philosophy, you could plant a bomb at the American Philosophical Association conference and kill off hundreds of philosophers at once. Before the police capture you, you could hop into the replicator. The person created by this process, according to the public norm, died by replication. The new person is not responsible for the crime, having just been born. Now this new person can apply for one of the many philosophy jobs that have suddenly become available.

If psychological continuity is the public norm, then neither Hatfield nor McCoy will have reason to commit the crime based on replication. As before, Hatfield will not because he will consider this equivalent to suicide. McCoy will not because the public norm says that McCoy will survive the replication and be subject to criminal sanctions. Since a public norm of somaticism is more likely to lead to negative social consequences, this gives us some reason to reject somaticism.

(There are other policy options we might explore, e.g., we could simply shoot anyone who is not a somaticist, or ban replicating technology. For some, these two policies might be very similar in terms of their practical effects. Imagine, similar to the stroke case, that without replication someone will die. Banning replication technology will end his or her life just as surely as being shot would. Obviously the two policies are not morally equivalent. Rather, it is to point out that banning such technology would come with some huge costs for some.)

7. No branching

Debates about identity preservation and uploading invariably get hung up on the “branching” problem, and this probably provides the strongest support for somaticism. The problem is that it seems there is only one of me. But uploading seems to allow the possibility that there could be hundreds, if not millions, of “me.” But if there can be only one of me, then uploading does not preserve my identity. It is clear how this problem arises given our previous discussion of the uploading process. Imagine my brain is scanned and the relevant information is recorded. Instead of being uploaded to a single computer with robotic body, imagine a thousand robot brains are encoded with the information. Of course it seems possible that thousands of robots could awaken in the same instant, all claiming to be Mark Walker. (And what a wonderful world this would be!)

Using the equivalence thesis we can see how this is exactly the same problem as the problem of branching that philosophers discuss in connection with carbon-to-carbon transfers. Parfit extends his Teletransporter case in exactly this way:

Several years pass, during which I am often Teletransported. I am now back in the cubicle, ready for another trip to Mars. But this time, when I press the green button, I do not lose consciousness. There is a whirring sound, then silence. I do not lose consciousness. I leave the cubicle, and say to the attendant: “It’s not working. What did I do wrong?”

“It’s working,” he replies, handing me a printed card. This reads: “The New Scanner records your blueprint without destroying your brain and your body. We hope that you will welcome the opportunities which this technical advance offers.” (Parfit 1987, 199)

Of course there is no reason to stop at one replica. Using Parfit’s Teletransporter thousands of organic molecule-for-molecule identical persons could awaken in the same instant, all claiming to be Mark Walker. (And what a wonderful world this would be!)

Notice that I did not say that any of the thousand persons claiming to be Mark Walker are me. Somaticism will deny that any of the thousand replicas are me; only the original is me. If the original is destroyed, and a thousand replicas are made, then somaticism will claim that I did not survive.

What does the psychological account have to say about multiple replicas? Here opinions differ. On the one hand, it seems that if there are multiple replicas, and they are all psychologically indistinguishable from the original, then each of them has as good a claim to be me, and so they are all me. The contrary “no-branching” view is that at most one replica is me, for there can be only one me (Shorter 1962).

The question then is whether there can be “branching”: more than one of me. I will argue that both sides of the debate are correct; there is a sense in which there can’t be more than one of me, and a sense in which there can be multiple versions of me. The first step in our argument is to get a little clearer about the no-branching argument, which may be schematized as follows (where “**P**” stands for “premise” and “**C**” for “conclusion”):

The No-branching Argument

P1: Multiple replicas X, Y, Z... of an individual O (the original) are numerically non-identical with each other, that is, X is not identical with Y or Z, Y is not identical with X or Z, and so on.

P2: Preservation of personal identity requires preservation of numerical identity.

C: Therefore, not all replicas X, Y, Z... preserve personal identity of O.⁷

It is worth distinguishing this argument from a similar but less serious objection. The less serious objection is that if there are a thousand replicas, then they will quickly have psychologically distinguishable properties. All thousand replicas will not fit in the same cab, for example, and so will have different experiences leaving the replicating center. Their psychological states will only diverge further over time. Riffing on Parfit's example, we can imagine a replica waking on Mars, the Moon, and Pluto all at the same time. Each will almost immediately have different experiences, and so quickly will be psychologically different.

Even if this is conceded, it does not answer the question of the status of the replicas at the moment they are created. Imagine the thousand replicas are all created at the same instant, and each awakes in a separate but identical room. At the instant of awakening, there will not be any psychological divergence⁸ and so the argument from diverging experience tells us nothing about the identity of the thousand replicas at this moment.⁹

I want to suggest that the problem with the no-branching argument is that there is a critical ambiguity. To explain the ambiguity it will be helpful to review the type/token distinction.

8. Types and tokens

The nineteenth century philosopher Charles Peirce is credited with first making the type/token distinction. Peirce's own example involving the individuation of words is as instructive as any:

A common mode of estimating the amount of matter in a manuscript or printed book is to count the number of words. There will ordinarily be about twenty *the*'s on a page, and of course they count as twenty words. In another sense of the word "word," however, there is but one word "the" in the English language. (Peirce 1906)

There are twenty tokens of the word *the*, but a single type of the word *the*. The argument to be canvassed is that if we think of personal identity as ambiguous between types and tokens, then the no-branching argument may be rejected.

We may approach the issue by recasting the previous argument by reference to a work of literature; so let us consider the no-branching argument applied to Shakespeare's *Hamlet*. We may reconstruct the argument first as about tokens, and then about types.

No-branching Token Argument

P1': Multiple replicas X, Y, Z... of an individual O (the original *Hamlet* penned in Shakespeare's hand) are numerically not (token) identical with each other, that is, X is not (token) identical with Y or Z, Y is not (token) identical with X or Z, and so on.

P2': Preservation of play-identity requires preservation of (token) numerical identity.

C': Therefore, not all replicas X, Y, Z... preserve play-identity of O.

It is pretty clear where this argument goes wrong: **P2'** is false. The original token of *Hamlet*, written in Shakespeare's hand on paper created over four hundred years ago, is now long lost. But the same play that Shakespeare wrote can be read today. The no-branching token argument fails.

No-branching Type Argument

P1'': Multiple replicas X, Y, Z... of an individual O (the original *Hamlet* penned in Shakespeare's hand) are numerically not (type) identical with each other, that is, X is not (type) identical with Y or Z, Y is not (type) identical with X or Z, and so on.

P2'': Preservation of play-identity requires preservation of (type) numerical identity.

C'': Therefore, not all replicas X, Y, Z... preserve play-identity of O.

It is pretty clear where this argument goes wrong: **P1''** is false. I may have bought my copy of *Hamlet* at a different bookstore than you, but still, we are reading the same play. At least in the case of plays, the type version of the no-branching argument fails. Thus, the no-branching argument, in both its token and type formulation, does not look the least bit plausible when applied to literature.

9. The type/token solution to personal identity

In this section I will say a little about the type/token (TT) account of personal identity¹⁰ and then see whether the no-branching argument has any traction against it. In the case of literature, the tokens of *Hamlet* are individuated according to the physical implementation: my *Hamlet* is in a different spatial location from your *Hamlet*. The *Hamlet* type is an abstract entity, which particular tokens of *Hamlet* embody. Similarly, the TT solution to personal identity says that tokens of a person type are individuated in terms of physical implementation: each replica will have a different spatial location. The person type is the abstract entity, which the various tokens are all embodiments of.

We previously rejected somaticism, but this is because we had yet to survey the type/token distinction. The version of the type/token view that we should adopt says that somaticism is correct about tokens, and the psychological account correct about types. Consider then the case where the original Mark Walker is scanned and destroyed and a thousand replicas are created. Somaticism, as a theory about tokens, says that the original token was destroyed, and a thousand new tokens created. The psychological account applied to types says that the Mark Walker type continues to exist, and indeed, is multiply instantiated.

The ontological status of abstract entities is a perplexing and contested issue (Wetzel 2009), but there is no reason to think that it is more perplexing in the case of persons rather than literature, and we are committed to types in the case of literature.¹¹

Can the non-branching argument be deployed against TT? Assuming that types can have more than one token, non-branchers cannot allow the notion of types to have a role in personal identity. So, to disambiguate the original non-branching argument, it must be about tokens:

The No-branching Argument in terms of Tokens

P1''': Multiple replicas X, Y, Z... of an individual O (the original) are numerically [token] non-identical with each other.

P2''': Preservation of personal identity requires preservation of numerical [token] identity.

C''': Therefore, not all replicas X, Y, Z... preserve personal identity of O.

There are two problems with this argument. First, it is question begging. The entire issue is whether personal identity can be explained in terms of preservation of type identity, and so **P2'''** prejudices the issue.¹²

The other problem is that it is difficult to see how one can insist on non-branching without collapsing into somaticism. To see this, consider the case where the original Mark Walker's body, O, is destroyed when

three replicas X, Y, and Z are created. Either O is not identical with any of X, Y, Z, or O is identical with one of X, Y, Z. If the former, then non-branching is simply somaticism in disguise. If it is asserted that O is identical with exactly one of X, Y, Z, then any choice would be arbitrary in the sense that choosing one among the thousand to be *The Mark Walker* would not be choosing based on any intrinsic differences. We could, for example, have all the replicas draw a number out of a hat and designate the winner of the lottery *The Mark Walker*. But an appeal to a lottery shows that precisely no intrinsic properties are used to individuate: it is the process (the lottery) that does the individuating. We could do the same for *Hamlet*. We could assign a number to every extant copy of *Hamlet* and have a lottery to find out which is *The Hamlet*, and which are mere copies. But, of course, no one would be impressed by this.¹³

Criticizing the non-branching argument is not a positive argument for TT, but it does suggest that TT need do little to prove itself more plausible than non-branching. However, in terms of a positive argument for TT, the fact that it provides a number of intuitively plausible consequences speaks in its favor:

- i. The TT solution explains why people might be reluctant to enter the Teletransporter in Parfit's original example. It requires sacrifice of their token identity, for little compensatory gain (three weeks in a space craft). In other words, it can explain how we might survive (as a type), even though we may regret some loss of identity (since the original token is now dead).
- ii. The TT solution explains why it would be rational to sacrifice one's token identity in the stroke case: the loss of token identity is not inconsiderable, but the replica is type identical, and the new token replica will be in better shape than the original token.
- iii. The TT solution explains why, in the Vorlon radiation case, we are not likely to feel much threat to our identity: the type survives destruction of each token, and each token is only a day old. Analogously, other things being equal, I would feel more upset about having my copy of *Hamlet* from my undergraduate days stolen, and less worried about a copy just acquired yesterday. In both cases, of course, the same play type is stolen.
- iv. The TT solution explains why in Parfit's modified transporter case – where a replica is created on Mars, but the original on Earth is not destroyed – the person on Earth has more claim to being Parfit than the replica on Mars. The person on Earth is both type and token identical with Parfit from the previous month (assuming he hasn't used the Teletransporter in the meantime). The replica on Mars is merely type identical.¹⁴
- v. The TT solution explains in a satisfactory manner what happens when the original is destroyed to make multiple replicas. Each replica is type identical with the original, but none of the replicas are token identical with the original. This avoids the embarrassment of having to say which of the indistinguishable replicas is identical with the original.

10. Should I upload?

I have tried to strike some compromise between saying that there is no loss of identity in replication (and by our equivalency thesis, uploading), and the position that survival is impossible. Still, it may look as though this is tantamount to an argument against uploading: if there is any loss in uploading, even if it is only token identity, why would anyone want to sacrifice some identity? The answer is that there are considerable advantages (or at least purported advantages) to being uploaded, including immortality and enhancement.

Except for the completely reckless, forgetful or lazy (ahem), everyone backs up his or her valuable computer files. But once we see that people too can be backed-up, it appears that virtual immortality is assured. For so long as there are operating computers, one can simply transfer the files that comprise oneself from computer to computer. If the hardware on one computer fails, you simply move to another computer. Suppose a piano falls on your robotic body. No problem. A new robotic body is brought out of the closet and a backup copy of you is uploaded. What formerly would have meant certain death is now only a small inconvenience. As Freeman Dyson long ago realized, the question of how long one might live quickly resolves to how long the universe will remain habitable (Dyson 1979), hence, the term “digital immortality” is sometimes used to refer to this prospect.

As for enhancement, one possibility is that our senses could be radically enhanced: robots presently make use of a sensory apparatus that detects light in parts of the spectrum not available to (unaided) human vision (e.g., infrared, x-rays, etc.), sounds that are beyond normal human auditory range, and so on. In terms of enhancing cognition consider that it is a relatively routine matter to add memory or computing power to today’s computers. If one is uploaded to a computer, then it seems that it would be a relatively routine matter to enhance one’s memory or cognition: just add more computer memory or processing power. The sky is literally the limit here. Anders Sandberg (1999) has done some preliminary calculations to suggest that planetary scale computers, “Jupiter-sized brains” might be possible. How powerful and how smart would such brains be? It is, obviously, hard to say. Certainly they would eclipse us by a greater margin than we eclipse the cognitive powers of your typical lab rat. Along with such enhanced cognition would come awesome powers to manipulate the physical world, for there is some truth to the saw that knowledge is power. In short, and without too much hyperbole, those who upload may well be on their way to godhood.¹⁵

It is beyond the scope of this paper to argue that these purported benefits of uploading really are benefits, but, if they are, the temptation to upload is clear. And just like in the stroke case, it is clear why it might be rational to forgo token identity survival for these advantages.

Acknowledgments

I would like to acknowledge the following persons for their assistance: Nick Agar, Russell Blackford, Jamie Bronstein, and Tim Cleveland.

Notes

1. Hans Moravec (1988) discusses ways around this challenge.
2. There are many problems for the bodily criterion of personal identity that I will not discuss here. See Williams (1973) and Olson (2006).
3. Parfit has found similar reactions to the case. He even concedes that he has some residual doubt that he would survive teletransportation (Parfit 1987, 279).
4. Moravec (1998, 117) suggests that this is like being attached to “jelly.”
5. From the television series: *Babylon 5*.
6. The protagonist and somaticist in John Perry’s fun little dialogue accepts death rather than ceding ground to the psychological account (Perry 1978).

7. See Williams (1973). Thomas Reid seems to have had a similar argument in mind: see Perry (2008) and Martin and Barresi's editors' introduction (Martin and Barresi 2003).
8. At least where psychological states are narrowly construed (Putnam 1981).
9. Even Nozick's (1981) closest continuer theory would not be able to choose one among the many.
10. Williams (1973, 80-81) considers something analogous to the present proposal.
11. This is not to argue, *à la* Quine, that we are committed to realism about types. Rather, it is to say that whatever one's preferred account, realist or nominalist, of types as applied to words or literature, the same account can be applied to persons. Nominalism is not the same as skepticism: nominalists can offer an account of types such that it makes sense to say that you and I are reading the same play, even if we own different copies of the play.
12. Patrick Hopkins has an argument that seems to show that the type construal is not possible: "The relationship of 'identity' is a very strong, and necessarily strong, concept that strictly refers to literal sameness – not similarity. When we use 'identical' in this sense, we are not saying that two things are just very like each other ('identical twins'), or even exactly similar ('identical cars'), we are saying that 'two' things are actually one and the same specific thing" (Hopkins n.d.). I'm not sure what Professor Hopkins would say about *Hamlet*: are he and I reading exactly similar plays or are we reading the same play? The former sounds perverse to these ears. But even granting this is the correct description, it seems that "exact similarity" is all the identity we may want in certain cases (such as the stroke case).
13. Indeed, we could solve every paradox – a set of statements that are individually equally plausible, but mutually inconsistent – by assigning a number to each statement and rejecting the loser of a lottery. Obviously, this is no way to solve a paradox.
14. Parfit (1987, 293-97) discusses the possibility of individuating in terms of types and tokens. He argues that it is the type that matters, not the token. As noted above, I disagree: both type and token are important.
15. A related objection is that radical enhancement will threaten identity (Walker 2008).

References

- Agar, N. 2010. *Humanity's end: Why we should reject radical enhancement*. Cambridge, MA: MIT Press.
- . 2011. Ray Kurzweil and uploading: Just say no! *Journal of Evolution and Technology* 22(1): 23-36.
- Dyson, F. 1979. Time without end: Physics and biology in an open universe. *Review of Modern Physics* 51(3): 447-60.
- Hopkins, P. n.d. The ghosts haunting transhumanism, or, why uploading will not work. *International Journal of Machine Consciousness*. (Pub. forthcoming.)
- Locke, J. 1975. *An essay concerning human understanding*. Oxford: Oxford University Press. (Orig. pub. 1690.)

- Martin, R., and J. Barresi. 2003. *Personal identity*. Oxford: Blackwell.
- Moravec, H. 1988. *Mind children: The future of robot and human intelligence*. Cambridge, MA: Harvard University Press.
- Nozick, R. 1981. *Philosophical explanations*. Oxford: Oxford University Press.
- Olson, E. 2002. Personal identity. *The Stanford encyclopedia of philosophy*. Available at: <http://plato.stanford.edu/entries/identity-personal/> (Orig. pub. August 2002; revisions October 2010.) Accessed November 15, 2011.
- . 2006. Is there a bodily criterion of personal identity? In F. MacBride, ed. *Identity and modality*. Oxford: Oxford University Press.
- Parfit, D. 1987. *Reasons and Persons*. Oxford: Oxford University Press. (Orig. pub. 1984.)
- Peirce, C. S. 1906. Prolegomena to an apology for pragmatism. *Monist* 16: 492-546.
- Perry, J. 1978. *A dialogue on personal identity and immortality*. Indianapolis, IL: Hackett Publishing Company.
- . 2008. *Personal identity* (ed.), Berkeley, CA: University of California Press.
- Putnam, H. 1981. *Reason, truth, and history*. New York: Cambridge University Press.
- Sandberg, A., 1999. The physics of information processing superobjects: Daily Life among the Jupiter Brains. *Journal of Evolution and Technology* 5(1).
- Sandberg, A. and N. Boström. 2008. *Whole brain emulation: A roadmap*. Future of Humanity Institute, Oxford University. Available at: <http://www.fhi.ox.ac.uk/Reports/2008-3.pdf> Accessed July 3, 2010.
- Searle, J. R. 1980. Minds, brains and programs. *Behavioral and Brain Sciences* 3: 417-57.
- Shorter, J.M. 1962. More about bodily continuity and personal identity. *Analysis* 22(4): 79-85.
- Walker, M. 2006. A moral paradox in the creation of artificial intelligence: Mary Poppins 3000s of the world unite! In *Human implications of human-robot Interaction: Papers from the AAAI workshop*, edited by Ted Metzler, California: AAAI Press, 2006: 23-28.
- . 2008. Cognitive enhancement and the identity objection. *Journal of Evolution and Technology* 18(1): 108-115.
- Wetzel, L. 2009. *Types and tokens: On abstract objects*. Cambridge, MA: MIT Press.
- Williams, B. 1973. *Problems of the self*. Cambridge: Cambridge University Press.

