

The Metaphysics of Uploading

(with Joe Corabi)

Journal of Consciousness Studies, forthcoming (Special Issue – Symposium on David Chalmers' paper on the singularity.)

Forthcoming as a reprint in *Uploaded Minds*, Russell Blackford, Wiley-Blackwell, 2014. (Includes our response to David Chalmers' reply to the present paper.)

Introduction

Metaphysics is a matter of life and death. When it comes to making a decision about whether to upload when the singularity hits, the devil is in the metaphysical details about the nature of substance and properties, or so we'll urge. An *upload* is a creature that has its thoughts and sensations transferred from a physiological basis in the brain to a computational basis in computer hardware. An upload could have a virtual or simulated body or even be downloaded into an android body – indeed, an upload could even primarily exist in the digital world, and merely occupy an android body when needed.

Imagine life as an upload. From a scheduling perspective, things are quite convenient: on Monday at 6PM, you could have an early dinner in Rome; by 7:30 PM, you could be sipping wine nestled in the hills of the Napa Valley; you need only rent a suitable android in each locale. Airports are a thing of the past for you. Bodily harm matters little – you just pay a fee to the rental company when your android is injured or destroyed. Formerly averse to risk, you find yourself skydiving for the first time in your life. You climb Everest. You think: if I continue to diligently backup, I can live forever. What a surprising route to immortality.

But wait! Metaphysics is not on your side. As we'll now explain, the philosophical case for surviving uploading is weak. When you upload, you are probably dying. Now, you may have other reasons to upload, besides personal convenience and surviving death. If the uploading technology is accurate, you'd be creating a psychological duplicate of yourself. And if you merely seek a mental duplicate – say, to carry out your earthly tasks -- then we have no metaphysical bones to pick with you. But to the extent that your decision is fueled by a

suspicion that *you* will be the one carrying out the earthly tasks, your philosophical footing is tenuous.

Why consider the philosophical case for uploading now? This is merely the realm of science fiction, you may think. But science fiction often foreshadows science fact. And uploading may turn out to be a case of such convergence. According to certain scientists and philosophers, such as Nick Bostrom, Anders Sandberg and Ray Kurzweil, technological developments in recent decades have rendered what used to be seen as far-fetched science fiction tales goals that may very well be attainable during the technological singularity.¹

David Chalmers' insightful discussion of the singularity treats topics which mainstream philosophy of mind has ignored, uploading being one of them. In this piece, we develop Chalmers' discussion of uploading. In particular, we illustrate that by attending to issues in contemporary metaphysics – more specifically, to issues involving the metaphysics of substance – we can learn that on the assumption that one of the leading theories of substance is correct, the different forms of uploading are not likely to enhance or preserve *us*. One caveat: We can't discuss all the views of the nature of substance today, so we will merely discuss the leading views. Our strategy is to inspire deeper metaphysical reflection on the idea that uploading is a form of survival, and to illustrate that survival is unlikely on the assumption that one of the leading theories of substance is in play.

Here's how we will proceed: Section One introduces metaphysical issues that are key to determining whether you could survive uploading; then, in Section Two we argue that it is plausible, given these background metaphysical issues, that you will not survive uploading. At best, a different person is present after you attempt to upload, one who is a psychological duplicate of you. Section Three concludes.

Section One: Uploading and Personal Identity

For many, the attractiveness of uploading would be lost, or at least significantly diminished, if it turns out that we can't really survive. A pressing philosophical issue, then, is to determine whether an upload is *the very same person* as the individual who existed prior to the upload. Accordingly, this will be the primary focus of our paper.

¹ See, for instance, a technical report published by Oxford's Future of Humanity Institute: Sandberg, A. & Bostrom, N. (2008).

We've noted that survival is not the only reason to upload, though – for instance, one may simply want a creature to carry out one's earthly business. In addition to such practical motivations to upload, there are other, more metaphysical, issues to consider. You may believe that although the upload is not literally you, there is a special relationship between you and your upload. Consider that sometimes the relationship between persons at different times is not as simple as a question of identity or total distinctness. When a human embryo (call it "Ally") splits into twin embryos, it may not seem correct to describe either of the new embryos as identical to Ally (i.e., as the very same person as Ally).² But, at the same time, many people do not think it seems correct to say that Ally has died and been replaced by two different people either. According to some, the intimate relation that the new embryos share to Ally deserves to be treated differently than the relationship had by two people who were conceived from a different sperm and egg pair. We might label this special sort of relationship "continuation," bearing in mind that we will reserve the expression "survival" for numerical identity.

This leads us to a second kind of question: if uploading of a particular sort does not preserve identity, is the upload at least a *continuation* of the original (in the technical sense of "continuation" employed above)? And there are more questions still. Even if someone in the future was neither you nor a continuation of you, that future person might preserve aspects of you that might make bringing about the existence of such a person of tremendous import to you. For instance, what if you were told that you were about to die, but for that a reasonable sum of money you could form a biological clone of yourself right now that would have copies all of the most important positive memories from your life, as well as a number of the character traits that you most highly value in yourself. Many people, when placed in the situation, would happily pay the fee for this clone. Why? One reason might be that the existence of such a person would comfort your friends and family. But it is pretty clear that this is not the only consideration that may push you in a "yes" direction anyway. It is reasonable to conjecture that part of what may be driving you to pay for the clone is a desire for your distinctive experiences and characteristics to be preserved in the future for their own sake, even if not by you or a continuation of you. In a similar vein, then, we can ask of uploads whether they preserve enough of what is valuable about ourselves (or at least enough of what we consider valuable) so that, even if they are not identical to us or continuations of us, there will still be much from our standpoint that commends us to

² Assume, for the purposes of illustration, that it is uncontroversial that embryos are people.

upload, especially if doing so does not threaten the quality of our lives in any way (i.e., the lives of *us* or at least those who are continuations of us).³

As we can see from the above discussion, there are three kinds of issues that must be addressed in considering personal identity for any upload case. First, we want to determine if the upload has preserved *numerical identity*. (I.e., is it the very same person as before the upload?) If the upload is not numerically identical to the original person, we can then ask if it is at least a *continuation* of the pre-upload person, in the way that one of the embryos above might be thought to be a continuation of Ally. And if the upload is not even a continuation of the person, we can then turn to the question of whether it preserves enough of what is valuable about the pre-upload person (or at least enough of what we or the pre-upload person consider to be valuable) for uploading to be an attractive proposition from the agent's standpoint.⁴

While we will weigh in concerning all these matters, as noted, our primary task is to determine whether an upload would genuinely be you – that is, whether numerical identity obtains. So let us now ask: how can we determine when an individual who attempts to upload is numerically identical to the uploaded being? Here, we believe that attending to the metaphysical conception of a person, and particularly, the metaphysics of substance, will prove fruitful.

The metaphysics of substance

The notion of a person is philosophically rich; for one thing, persons are traditionally considered the bearers of rights, or at least entities that demand consideration in the utilitarian calculus. For

3. These separate issues are distinguished in Chalmers' paper on the singularity, albeit sometimes implicitly.

4. Strictly speaking, this question of whether the upload preserves what is valuable may be independent of questions of survival and numerical identity. For instance, depending on our specific views about personal identity, we could imagine an uploading procedure that preserves numerical identity, but fails to preserve much of what is valuable about the pre-upload person. (Perhaps it changes large numbers of character traits or erases all memory.) To keep things manageable, though, we will generally consider the question of preservation of what is valuable only when there is good reason to think that the upload is neither identical to nor a continuation of the pre-upload person. This will make sense in our investigation because, due to the nature of uploading, we will be hard-pressed to find a feasible form of uploading that has a decent claim to preserve numerical identity without preserving the various psychological features of a person that we would usually think of as the valuable ones.

another, understanding the person is central to our self understanding -- for instance, it helps us to understand what it is to be human, and what it is to be a reflective, conscious being. Further notice that there are metaphysical dimensions of the notion of a person: a person is a metaphysical object of some sort. For one thing, it is an entity that has a variety of mental features (or properties): *being rational, having certain kinds of conscious experiences*, and so on. These features are *mental properties* of the person. This naturally leads many metaphysicians who consider the notion of a person to take the position that the person is a *substance* – that is, an entity that continues to exist for at least a short period of time and is the bearer of properties. One key philosophical issue is whether persons are material or physical substances, (as the physicalist purports), or whether persons are distinct from physical objects, say, because persons (and arguably, nonhuman animals) are conscious or have souls.

Substances have some of their properties of essentially – that is, they require them for their continued existence. Others, they only possess contingently. So, for example, a person arguably has the basic capacity for rational thought essentially; in contrast, properties like the person’s particular weight and hair color are not essential, but contingent – after all, you could survive if you had a different hair color or weighed a bit more. Different theories of personal identity hold different properties to be essential to the nature of the person. Consider the leading theories:

1. **Soul theories** – this family of views holds that your essential property is that you have a soul or immaterial mind, where the soul and immaterial mind are non-physical substances that are distinct from any physical thing, such as the brain.
2. **Psychological continuity theories** – in their most general form this family of theories holds that you are essentially your memories and ability to reflect on yourself (Locke 1689; see Olson, 2010 and Perry, 1975). Today, certain proponents of uploading, such as Ray Kurzweil and Nick Bostrom, are broadly sympathetic to this type of view; they view the person as being essentially her overall psychological configuration, where one’s psychological configuration is to be detailed by a completed computational theory of the brain. (Kurzweil 1999, Bostrom 2003, Schneider 2009a)
3. **Materialism (or “physicalism”) about the person** – one is essentially the material that that one is made out of: the collection of molecules that makes up one’s brain, and arguably, the rest of one’s body.⁵

⁵ For a more extensive survey of the positions on personal identity see Eric Olson, 2010.

Each of these views has been framed by numerous individuals, and the details can differ in important ways (consider, for instance, the range of positions on the nature of the soul within Christianity). An important subtlety: While many psychological continuity theorists hold the view that everything is ultimately physical, (2) does not entail (3). What is important, and what makes (2) warrant being considered a distinctive position on its own, rather than being a version of (3), is that proponents of (2) contend that personal identity over time is a matter of psychological continuity, and not, in general, a matter of the person's physical or biological substrate.⁶ Further, psychological continuity can in principle be distinctive to personhood even if it turns out that persons, or other sorts of things, are not physical.

Some have seen fit to repudiate the reality of the person altogether:

4. **The No Self View.** According to the “no self view,” the self is an illusion. The “I” is a grammatical fiction (Nietzsche). There are bundles of impressions but no underlying self (Hume). There is no survival because there is no person (Buddha).

Upon reflection, each of these views has its own implications about whether one should upload. If you hold (1), then your decision to upload depends on whether you believe the upload would retain your soul or immaterial mind. If you believe (3), then uploading will not be a form of survival, because survival requires the same material substrate, and uploading changes it.⁷ In contrast, according to (2), uploading may be safe, because although it alters your material substrate it preserves your psychological configuration. Finally, (4) contrasts sharply with (1)–(3). If you hold (4), then the survival of the person is not an issue, for there is no person to begin with. You may strive to upload in any case, to the extent that you find intrinsic value in having an upload carry you're your earthly tasks, for instance.

⁶ In practice, most psychological continuity theorists have not addressed the metaphysics of substance and taken a position on the relation of personal identity questions to metaphysical issues about substance. In spite of their lack of interest, though, these theorists' views may commit them to the claim that persistence of substance matters to personal identity in a direct way. The reason is that they often believe that it is important that memories be preserved, and the continuation of real memories (as opposed to memory copies) is reasonably linked to specific substances.

⁷ Later, we'll consider a form of materialism that holds uploading is a form of survival, but the issues are complex; here, we have stated a simple formulation of materialism.

One common thread that underlies all of the views of persons (with the exception of the no self view, of course), is that advocates of these views tend to hold that the person or self is some kind of substance – an object-like entity that has properties and continues to exist for some period of time. Now, some have repudiated the category of substance; some, for instance, instead accept an ontology of events or believe that an ontology of properties can secure the rejection of the category of substance. But these views are implausible, we believe, and one of us has treated these issues elsewhere (Schneider 2011).

So, given the centrality of the metaphysical category of substance to debates on the person, it will be useful to consider some particular views of the nature of substance.

Leading Conceptions of Substance

What is the nature of substance? Contemporary metaphysics tends to focus on two leading theories: the bundle theory and the substratum theory (Armstrong 1989; Loux 2002; Schneider 2011). Both approaches are similar insofar as they hold that substances are not metaphysically basic; instead, the category of substance reduces to (*inter alia*) properties, that is, the features of things (e.g., the blueness of the sky, the mass of a neutrino). The idea that object's natures involve their properties is appealing, so it is not surprising that both the substratum and bundle theories have a distinguished history and are currently so well received. For when we conceive of an object what comes to mind are its features. But beyond this point of agreement, there are important differences between the two theories.

According to the **Bundle Theory**, substances are bundles of the properties they possess. Of course, not every bundle of properties is an object, so we should ask: what unites bundles that are really substances? Here, the bundle theorist suggests a relation called “compresence” (or “co-instantiation”, “togetherness”, “collocation”), where the compresence relation is usually taken as primitive. The **Substratum Theory**, in contrast, holds that objects' natures are not exhausted by their properties. Over and above their properties, substances have substrata, a core that bears properties but is not itself a property. As our sensory access to objects is through their properties, substrata will seem rather mysterious to you: indeed, Locke, who was himself a substratum theorist, comments that they are something “I know not what” (1689, II, xxiii, §2). So why should we believe that objects have substrata? Suffice it to say that belief in substrata comes from appreciating deficiencies in the bundle theory; leading objections include that the compresence relation has been notoriously difficult to spell out and that the bundle theory turns

objects into properties (Goodman 1966; Armstrong 1989, 70–72, 1978, Ch.9, Sec. IV; Russell 1948, 312; Schneider 2011).

In addition to all this, it is also worth emphasizing that key to debates over the nature of a person is the matter of whether person's nature is physical or nonphysical. Are persons identical to their bodies, as (3) purports, or is there something more to their nature, something nonphysical, that outruns anything physical, as (1) contends? Here, to the extent that one endorses substances to begin with, philosophers tend to be either *substance physicalists*, holding that the self or person is ultimately a physical substance, or they reject physicalism for a *substance dualist* view in which the self or person is a nonphysical substance.

Now, in keeping with the aforementioned distinction between the bundle and substratum theories we can introduce different kinds of substance dualism. According to the **substratum version of substance dualism** (or “SSD”) persons are ultimately substrata instantiating only mental properties, not physical/functional ones. What we might pretheoretically describe as the physical parts of such persons are really entirely different substances that are not part of persons at all. There are separate substrata having separate bunches of physical properties. These physical substances presumably then causally interact with persons.

The substance dualist bundle theory of substance (“SDBT”) says that persons are ultimately bundles of just mental properties. Parallel to what we saw with the substratum version of substance dualism, the purportedly physical parts of such persons are separate (purely physical) property bundles, and not really parts of the persons at all.⁸

We can now appreciate that although our discussion is by no means exhaustive there are several views about the nature of the person as substance that are possible.

A further issue is what the persistence of an object, or, more specifically, a person, consists in. Two views are widely defended in the contemporary literature. Let us turn to this matter.

The persistence of persons

8. Although we will not deal with them explicitly, views that hold that persons are composites of two substances will fall within the scope of arguments we give later. Because they raise no new philosophical issues and addressing them requires us to introduce significant peripheral complication, we omit direct treatment. We also will not have an opportunity to treat hybrid substances (see Schneider 2011).

The two leading views of the persistence of objects are **three-dimensionalism** (often called “endurantism” or simply “3D-ism”) and **four-dimensionalism** (often called “perdurantism”, “the doctrine of temporal parts” or “4D-ism”).⁹ **Three-dimensionalism** claims (roughly) that objects persist by enduring or “sweeping” through time, remaining wholly present at each time during which they persist. **Four-dimensionalism**, on the other hand, claims that objects persist by “perduring” or being “spread out” through time, existing at particular times by containing temporal parts that are present at those times. So, for instance, if we assume that your desk is an object that persists through the interval between 1PM and 2PM on January 10, 2012, a three-dimensionalist will claim that it persists because there is an object—namely, the desk—that exists at the beginning of that interval, the end of the interval and throughout the interval.¹ The four-dimensionalist, on the other hand, holds that it persists because there are a bunch of instantaneous objects—objects we might intuitively identify with your desk *at a particular instant*—and there is such an object for every instant of time during the interval. The key is that these objects are part of a set or sum of such objects (called temporal parts) that together compose your desk.

3D and 4D views can be occupied by proponents of either the bundle or substratum theories. Consider the 4D view, for example; as discussed, at a given moment a time slice of you, not you in your entirety, exists. This time slice is a bundle, as per the bundle theory. But there are bundle-bundle views of substance, according to which (in broad strokes) the entire spacetime worm that is you is a kind of bundle of the time slices. To keep the discussion simple, we will speak in terms of the 3D view, noting particular issues arising in the context of the 4D view in the final section of the paper.

⁹ Although one could certainly hold mixed views (whereby, for example, some objects endure but persons perdure), but we will always interpret three-dimensionalism and four-dimensionalism to be views specifically about the persistence of persons. For prominent statements of three-dimensionalism (as a general metaphysical view), see Baker (2000), Chisholm (1976), Haslanger (1994, 1989a, 1989b), Merricks (1999), Thomson (1998), Van Inwagen (1990), Zimmerman (1999, 1998a, 1998b), and Wiggins (1980). For prominent statements of four-dimensionalism, see Armstrong (1980), Carnap (1967), Goodman (1951, ch. IV, sect. 1), Lewis (1986, 202-204), Quine (1960, sect 36), Sider (2001), Smart (1972). The current four-dimensionalist view has historical antecedents in Hume (1739/1978, book I, part 4, sects. 2,6) and Jonathan Edwards (1758, part 4, ch. 2).

Again, this discussion of views of the metaphysics of the person is not exhaustive, and we can see that it is not even possible to treat in depth the positions we raise. But in the spirit of illustrating the import of the metaphysics of the person to the question of whether an upload would genuinely be you, we shall now turn to a more detailed discussion of whether, according to certain paradigmatic positions, the upload would really be you.

Section Two: Putting the Metaphysics to Work

Instantaneous, Destructive Uploading Attempts

Now let us ask: Could you survive destructive instantaneous uploading? Recall that this is a form of uploading in which all at once, all the precise information about your mental functioning is measured, and your brain is destroyed. All of the information is transmitted to a computer host at some distance away. Let us first consider this question in the context of the substratum view; again, according to the substratum view, a person is a propertied substratum that endures only if the substratum endures. On the assumption that persons are indeed propertied substrata, is it plausible that you would survive destructive instantaneous uploading?

We doubt it. Consider first that the mechanism by which the person (i.e., the propertied substratum) would move instantaneously from the brain to the computer is problematic, even on the assumption that only a short distance needs to be traversed. Not only does this involve an unprecedentedly rapid kind of motion for a person to follow, but this sort of motion is oddly discontinuous. For it is not as though the person moves, little by little, to the computer, so that a step-by-step spatial transition from brain to computer can be traced. Since information is being uploaded, the information has to be processed and reassembled in the computer host before anything like a functional duplicate of the original brain can be obtained. Hence, the person exists at one moment in a brain, and then ceases to exist for a brief period while the information is being transported, and then comes back into existence in the computer at some distance away.¹⁰

While it is possible that substrata behave in such bizarre ways the issue is whether it is justifiable to believe that they really do so. Ordinary physical objects simply do not behave like this. Substratum theorists hold that substrata underlie ordinary objects, after all, and this sort of spatial and temporal discontinuity is incompatible with standard views about the endurance conditions

10. Some adherents of this general position will reject the claim that people have spatial locations at all—namely, full-blooded Cartesian dualists. (Descartes himself famously claimed that people—i.e., souls—have no spatial locations.) Although the arguments we give here will not come into direct dialectical contact with the full-blooded Cartesian view, presumably full-blooded Cartesians have intuitions about what sorts of physical entities can causally interact with souls. They will likely endorse constraints on what kinds of bodies may interact with a particular soul that allow parallel arguments to be constructed against the claim that uploading preserves identity on a Cartesian view. For example, most Cartesians will probably endorse the claim that a particular Cartesian soul causally interacting with a biological body at one time could not begin causally interacting with a computer-based body 1,000 miles away from that biological body a fraction of a second later.

of ordinary objects—these intuitions are much stronger than any particular intuitions about the continued existence of a person in this sort of scenario.

Second, unless the brain is truly destroyed at the *very* instant the information is obtained from it it will be the case that for at least a small interval the person will continue to exist at the location of the brain even after the information has been transmitted to the computer. If the computer is fast and near enough, and it takes long enough for the brain to be destroyed, this could result in the person remaining where the brain is even after the information has begun its journey to the computer (or perhaps after the information has finished its journey there).¹¹ Only later does the person “catch up” to the information and come to be located where the computer is. Once again, the strangeness of this sort of behavior violates strong intuitions about the endurance conditions for particulars, and substrata are particulars. Consequently, as a result of both of these considerations, it is sensible to conclude that identity will not be preserved.

You may further think that the following is evidence against the survival of the substratum: the upload presumably instantiates various silicon-based properties as opposed to carbon-based ones. But this point will not be well received by proponents of the aforementioned psychological continuity view. On this view, it is not the underlying physical material that is essential to one’s being the same person over time; what matters is that the upload be psychologically continuous with the pre-uploaded person. But notice that even if one were to grant that one’s underlying physical features could differ this radically before and after uploading, on the view we are considering, the continuity theorist is also a substratum theorist, and we believe that she would lack justification for claiming that destructive instantaneous uploading is a form of survival. Even if the upload has all the same mental properties as the original person, this would not entail that the substratum was the same, and this is required for numerical identity to obtain, (i.e., for the person to survive). The problem is that substrata are particulars, and, as we’ve just observed, it is questionable that particulars behave in the bizarre manner that would be required for uploading to really be a form of survival. It is more sensible to hold that the uploaded being is merely a psychological copy of the original, having a different substratum. *The interesting upshot is that while we normally believe that the psychological continuity view is friendly to uploading, considering it in light of the substratum view suggests otherwise.*

11. If the person does not “catch up” until after the computer has reassembled the information and created a conscious being, then what we have is a scenario that is best classified as non-destructive uploading, albeit a non-destructive uploading situation where the original only lasts a short while after the upload. These cases must be dealt with separately.

Turning to the matter of whether the upload is at least a continuation of the original person, (again, we have in mind the aforementioned technical sense of “continuation”), there is again reason for pessimism. The paradigmatic instances of continuation are ones where there is physical continuity between the bodies of the survivors and that of their predecessor. Consider, for instance, stock examples like the identical twin scenario from earlier, or transplants of one cerebral hemisphere into a new body. In cases like these, the claim to survival is based strongly on the physical continuity between the bodies of continuations (and more specifically, brain material in the brain transplant case) and the body of their predecessor. So, insofar as we have any intuitions about survival, the intuitions track this sort of continuity. But, as we have already seen, the relevant continuity is utterly lacking in uploading, since the “birth” of the upload is spatiotemporally separated from the “death” of the original individual, and moreover the “body” of the upload doesn’t even possess the same broad-ranging kinds of physical properties as the body of the original person – that is, the pre-upload is a carbon-based being in which mental properties are in a cellular substrate, whereas the upload is not.

Now that we have considered destructive uploading in the context of the substratum view of substance, we are in a position to see that the pessimistic conclusions of this section do not really depend on either the substratum or 3D views, nor even upon whether one is a materialist or a substance dualist. *For the main reasons for being pessimistic have to do with reservations about positions that attribute strange and discontinuous motions of persons (in contrast to the behavior of other macroscopic substances) in order to claim that uploading is compatible with survival.*

This being said, we still plan to consider uploading in the context of the bundle theory of substance in more detail, so let us turn to this matter. Here, there is an additional problem for 3D bundle theories that doesn’t arise for parallel substratum versions. With substratum versions of the 3D view, the substrata, together with the essential properties, individuate persons and provide the particularity that is required for multiple persons to exist and be distinct from one another. But with bundle versions, however, only the properties themselves can provide the needed particularity. Now since, according to 3Dism, persons are to be identified with the particulars that bear what we would intuitively think of as *their* (i.e., the persons’) properties, 3D versions of the bundle form of the view must claim that the bundle endures. But it seems like the only way for a bundle to endure would be via the endurance of all the essential properties in the bundle (as well as their continued co-presence with other properties in the bundle). Once we acknowledge this, though, it becomes difficult to see how any form of destructive uploading will achieve the desired preservation of the bundle in the new location. The new physical body will not be

identical to the old one; after all, it will be in a completely new location and be composed of silicon building blocks rather than carbon based ones. And if the physical body of the person in the new location is completely different, it is very difficult to believe that the properties in the bundle will have endured: even dualist views that do not require physical building blocks to be preserved would require the physical bundles that are supposed to causally interact with the nonphysical bundle (i.e., the soul or immaterial mind) to endure in spite of spatiotemporal interruption, and to causally interact with completely different physical properties before and after the temporal break.

Things don't get much better when we consider 4D views. With 4Dism, there may be a bit more flexibility in people's intuitions, since many people may feel that there are fewer constraints on what temporal parts could form an object over time than on how a single enduring substratum or bundle could behave. But even so, any additional breathing room is limited. Most 4D theorists about persons (as well as 4D views about ordinary physical objects) do believe that the spatiotemporal continuity of temporal parts is an important necessary condition for those things to be parts of the same object over time, at least so long as we are talking about "adjacent" temporal parts, and the object we are talking about is a non-gerrymandered one, as persons presumably are.

And, to continue a familiar theme, one's specific view of whether persons are physical or non-physical substances doesn't change things much either. Clearly, physicalists are in the worst shape, since the physical differences between the pre-upload person and the upload are most dramatic—they are made of different physical building blocks, after all, and these building blocks are obviously spatiotemporally discontinuous with the original building blocks. But substance dualists are not much better off, since many substance dualists believe that persons are essentially connected to their physical bodies, and even the ones that don't (i.e., those substance dualists who are more Cartesian) would have a hard time explaining how the soul would manage to "move" from the location of the original body to the location of the upload. (This motion could be literal or it could involve changing what physical objects the soul directly causally interacts with.)

To end the section on a more positive note, what about using uploading to preserve various characteristics that are, or are at least considered to be, valuable, such as character traits and qualitative replicas of memories? This is a difficult issue; sorting through it requires a more systematic investigation of people's intuitions and the various valuable characteristics. But here we see the greatest room for optimism: if people care a great deal about preserving copies of

their “memories,” character traits, and thought patterns, then uploading of this sort may be a feasible way to give them what they are looking for. But the extent to which this sort of thing matters to people remains a topic for further investigation.¹²

Gradual Destructive Uploading

Now consider the case of gradual destructive uploading. Imagine, for instance, a case like the one Chalmers presents, where a person’s brain is slowly replaced, neuron by neuron, by functionally isomorphic silicon units that transmit the functional information from the neurons just as they destroy and replace them.

You may suspect that this case offers better promise for survival than the instantaneous uploading case we considered earlier, but the change from instant to gradual uploading does not seem drastic enough to provide much additional reason for optimism. For one thing, although the transmission of information happens in a more piecemeal fashion than it does with the case of instantaneous uploading, there will nevertheless fail to be a functional isomorph until all the data from the original is uploaded. Thus, regardless of how long the overall replacement process takes or how small the incremental replacements are, there will still be a dramatic moment at which the data is assembled by the computer host and the isomorph is born. When we consider what happens at this moment, all of the old issues reemerge: e.g., if the substratum view is in play, does the person’s substratum instantly (or almost instantly) travel from brain to computer? Is there a temporal discontinuity between the person’s disappearance and reappearance? It is hard to believe that the same kinds of strong intuitions about endurance we saw before will not be violated once again. (And even if the upload is assembled gradually, there will still be an enormous spatial discontinuity between the person in the computer and the one in the brain.)

Interestingly, Chalmers seems sympathetic to the preservation of identity in the gradual uploading case. We find his reasons unpersuasive, though (assuming they are indeed intended as reasons to believe that numerical identity is being preserved). One consideration that he raises is that “[t] will be very unnatural for most people to believe that their friends and families are being killed by the process [of gradual uploading].” (Chalmers, 2012, p. 45). This is likely to be an untrustworthy folk intuition even if Chalmers is correct that it would be present; after all, attending to the sorts of subtle empirical and metaphysical details of cases like this, and attuning

¹² Parfit (1984), along the literature it has spawned, provides the classic foundation for much of this investigation.

philosophical intuitions to the relevance of these considerations, is not generally a strength of those without philosophical training.

Another reason that he provides is his confidence that consciousness will be continuous between the pre-uploaded individual and the uploaded one, and that the psychological continuity that results is “an extremely strong basis for asserting continuation of a person (p. 45).” If, on the one hand, this continuity requires the numerical identity of persons, then it seems uninformative. (I.e., it is uninformative if the continuity in question amounts to endurance of the very same stream of consciousness.) This is because, in order to know whether the continuity holds in the first place, we will need to know if numerical identity is preserved, and this is just what we are trying to get a handle on.¹³ If, on the other hand, all that is required for this continuity is that later mental states be caused by or be qualitatively similar to earlier ones (e.g., that qualitatively similar thoughts be entertained, or that later sensory “memories” be qualitatively similar to the original experiences), then plainly the sort of continuity in question could be shared by numerous individuals at later times; what if the information were sent to two computers instead of just one, after all? This allows us to form the following argument against continuity of consciousness (understood in a way that emphasizes causation or qualitative similarity) being anything near an epistemically sufficient condition for identity in these sorts of uploading scenarios:

- (A) If uploading preserves the continuity of consciousness, then the continuity of consciousness can be duplicated in multiple locations (since, after all, uploaded information can be sent to many distinct computers).
- (B) Uploads in distinct computer systems are distinct from each other.
- (C) If the continuity of consciousness can be duplicated in multiple locations and if uploads in different locations would be distinct, then, at most, one of the resulting

¹³ One might object that someone could verify, from the inside, that the same stream of consciousness has endured, and so verify that the same person has endured. But this approach faces a dilemma. Either the individual would need to use memory of her past thoughts, or not, as part of the verification process. If memory were involved, then there is the possibility that the individual could be misled, because even if the stream had not endured the feel from the inside would be the same as if it had endured. And if it is possible for the individual to be misled, then there can be no verification from the inside. If, on the other hand, memory were not involved, then the verification process would be entirely mysterious. It would require a person to have perfectly reliable beliefs about the past without relying on memory. But what faculty could possibly allow someone to have perfectly reliable beliefs about the past without using memory?

uploads is identical to the original person.

(D) One of the resulting uploads is identical to the original person. (This is an assumption that must be true in order for continuity of consciousness to ensure identity.)

(E) If one of the resulting uploads is identical to the original person, and if uploads in different locations are distinct, then whatever makes it the case that a particular upload *U* is identical to the original is non-causally dependent on events having nothing to do with the intrinsic characteristics of *U* or *U*'s substratum. (It could depend, for instance, on there not actually being other uploads, or on *U*'s original biological body being destroyed.)

(F) Whatever makes it the case that *U* is identical to the earlier person cannot be non-causally dependent on extrinsic factors (i.e., things having nothing to do with the intrinsic characteristics of *U* or *U*'s substratum).

From (A)-(F), basically by a succession of *modus tollens*:

(G) Uploading does not preserve the continuity of consciousness.

But this contradicts the obvious:

(H) Uploading does preserve the continuity of consciousness.

The only assumption that can reasonably be thrown out here is (D)—we cannot conclude that one of the uploads is in fact identical to the original. A *prima facie* attractive candidate for rejection might be (F), but this will not work. According to the 3D picture of persons, persons are fundamental entities that endure because the same underlying particular endures. Whether this particular endures, then, can hardly be non-causally impacted by things going on outside it, any more than the existence of a particular quark could be non-causally impacted by things going on outside *it*.

So much for Chalmers' tentative defense of the idea that gradual uploading preserves identity. Are uploads at least continuations? Again, this question may not be as clear cut as the parallel question for instantaneous destructive uploading, but the reasons for preferring a different conclusion are not particularly strong. There are still the same spatiotemporal discontinuity issues, and the fact that the biological brain is destroyed gradually and the information transmitted slowly doesn't seem to matter a lot.

Any form of non-destructive uploading is not likely to preserve identity, since the upload will clearly have no claim to be the person when the original brain is still very much in operation and

supporting consciousness and thought. (And, in addition, we still have all the same problems we discussed before in the context of destructive uploading.) Whether non-destructive uploads are continuations depends once again on whether being a survivor is compatible with the kind of spatiotemporal discontinuity we have seen. There is also a special issue that arises with non-destructive uploads—namely, whether continuation is compatible with the continued existence of the original person. All of the paradigmatic instances of continuation (e.g., the case of identical twins and the case of transplantation of hemispheres of the brain) seem to involve the original person ceasing to exist and being replaced by persons with an equal claim to intimate causal and substantial connection to the original. But when the original person clearly continues to exist, it is even more difficult than before to make the case that a mere qualitative duplicate with only distant causal connections to the original has any claim to being a continuation.

Just as we saw in connection with instantaneous destructive uploading, our conclusions about gradual destructive uploading don't depend strongly on embracing any particular theory of the nature of substance, of persistence, or of whether persons are physical or non-physical. But considering the issues in the context of a particular metaphysical positions, such as the substratum theory, provided a means of appreciating the rationale for a more pessimistic view of uploading.

Section Three: Conclusion

Now that we have discussed these two kinds of uploading scenarios, let us summarize where things stand. First, there is considerable reason to be pessimistic about instantaneous destructive uploading's ability to preserve identity or to produce continuations of the original person. For this would require odd spatiotemporal discontinuities, and such conflict with what we know of the general behavior of macrophysical objects. Second, there is also good reason to be pessimistic about gradual destructive uploading's ability to preserve identity or produce continuations, since exactly the same issues arise in the context of gradual uploading. Third, we saw that all forms of non-destructive uploading are unlikely to preserve identity, and not appreciably more likely to produce continuations. Finally, we saw that there was more room for optimism about whether uploading of all the various sorts preserves psychological aspects of persons that we deem worth caring about, although we noted that a more thorough investigation of people's intuitions on this topic is needed.

Sources Cited

Armstrong, DM (1989). *Universals: an Opinionated Introduction*. Westview Press.

Bostrom, N. (2003). "The Transhumanist Frequently Asked Questions: version 2.1." World Transhumanist Association. Retrieved from <http://transhumanism.org/index.php/WTA/faq/>.

Davidson, Donald (1980). *Essays on Actions and Events*. New York: Oxford University Press, 1980.

——— (1985). "Reply to Quine on Events," in *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. eds. Lepore, E. and B. McLaughlin. Oxford: Basil Blackwell, pp. 172–176, 1985.

Goodman, Nelson (1966). *The Structure of Appearance*. 2nd ed., Bobbs-Merrill.

Kurzweil, R. (1999). *The age of spiritual machines: When computers exceed human intelligence*. New York: Viking. Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. New York: Viking.

Locke, John (1689). *An Essay Concerning Human Understanding*.

Loux, M. (2009). *Metaphysics: A Contemporary Introduction*. New York: Routledge.

Olson, Eric T., "Personal Identity", *The Stanford Encyclopedia of Philosophy (Winter 2010 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2010/entries/identity-personal/>.

Perry, John. (ed.), 1975, *Personal Identity*, Berkeley: University of California Press

Russell, Bertrand (1948). *Human Knowledge: Its Scope and Limits*. Allen and Unwin.

Sandberg, A. & Bostrom, N. (2008): *Whole Brain Emulation: A Roadmap*, Technical Report #2008-3, Future of Humanity Institute, Oxford University URL: www.fhi.ox.ac.uk/reports/2008-3.pdf

Schneider, S., Nonreductive Physicalism and the Mind Problem, forthcoming in *Nous*.

-----(2009a) “Mindscan: Transcending and Enhancing the Brain,” in *Science Fiction and Philosophy*, (Susan Schneider, editor). Wiley-Blackwell, 2009. Reprinted in *Neuroscience and Neuroethics: Issues At the Intersection of Mind, Meanings and Morality*, Giordano J. (ed.) Cambridge: Cambridge University Press (2011).

-----(2009b) *Science Fiction and Philosophy*, (Susan Schneider, editor). Wiley-Blackwell, 2009.

Sider, T., 2001. *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press.