

Selective Optimism about Mind-Uploading

Clas Weber

The University of Western Australia

Abstract. Optimists about mind-uploading believe that we can survive uploading. Pessimists about mind-uploading, on the other hand, believe that we cannot survive uploading. An under-explored middle ground between the two is a selective form of optimism, which claims that we can survive some forms of uploading, such as gradual replacement uploading, but not others, such as scan-and-copy uploading. Is selective optimism about uploading a rational stance? In this paper I argue that the answer is: yes. The paper has a negative and a positive part. First, I defuse a recent objection against selective optimism from Wiley and Koene (2016). Wiley and Koene argue that gradual replacement uploading is metaphysically equivalent to scan-and-copy uploading, and hence optimism about the former and pessimism about the latter is unfounded. I show that Wiley and Koene's case for the supposed metaphysical equivalence fails, since there is a significant metaphysical contrast between the two types of uploading involving immanent causation. In the second part of the paper, I present a positive case for selective optimism about gradual uploading building on this contrast in immanent causation.

1. Introduction

Mind-uploading is the process of transitioning a person from their biological hardware to an artificial hardware. In gradual replacement uploading, this is attempted by incrementally replacing the biological neurons of a person's brain with artificial, functionally equivalent silicon chips. In scan-and-copy uploading, the brain is scanned, and the obtained information is used to create a simulation of the subject's brain on a computer. Optimists about mind-uploading believe that it is possible to survive mind-uploading. If the uploading process is successful, the upload genuinely is *you*. Pessimists about uploading believe that surviving uploading is impossible.¹ The best you can hope for is to create a copy rather than a genuine future version of yourself. An under-explored middle ground between the two is *selective optimism* applied only to certain versions of uploading but not to others. For instance,

¹ In contrast to Wiley and Koene, I do not distinguish between *person P survives* and *person P's identity is preserved*. It seems plausible that P survives from t_n to t_{n+1} if and only if there is a person P^* existing at t_{n+1} such that $P = P^*$. Put simply, you survive just in case there's a future subject that is you. Equally, P's identity is preserved from t_n to t_{n+1} if and only if there is a person P^* existing at t_{n+1} such that $P = P^*$. Put simply, your identity is preserved just in case there's a future subject that is you. On this interpretation, saying that one survives and saying that one's identity is preserved are simply different ways of saying the same thing.

it could be that we are able to survive the process of gradual uploading by neural replacement, yet unable to survive the process of destructive scan-and-copy uploading (for this taxonomy see (Chalmers, 2010; Wiley, 2014)).² A potential proponent of this stance may be David Chalmers who writes: “[...] I think that gradual uploading is certainly the safest method of uploading.” (Chalmers, 2010, p. 62). But does it really make sense to take different attitudes towards gradual replacement uploading vs. scan-and-copy uploading?

Wiley and Koene (2016) have recently presented an intriguing objection against selective optimism, arguing that we have to either be optimists or pessimists about both gradual replacement uploading and scan-and-copy uploading (Wiley and Koene themselves are optimists, but this commitment is not central to their case.) Their objection uses *instant* replacement uploading as a bridge linking gradual replacement and scan-and-copy uploading. In instant replacement uploading, all neurons are replaced in one fell swoop with artificial neurons. In a first step, they link scan-and-copy uploading and instant replacement uploading, claiming that the two are metaphysically equivalent. In the second step, they aim to establish an equivalence between gradual replacement and instant replacement uploading. In the final step, they conclude that, by transitivity, gradual replacement uploading and scan-and-copy uploading are equivalent, too. If true, there is no relevant metaphysical difference between gradual replacement and scan-and-copy uploading that could ground differences in personal identity. Thinking that it is possible to survive one but not the other would be baseless. They write: “both [forms of uploading] should actually be considered metaphysically equivalent in the ultimate identity status of the minds they produce, either successful preservation of personal identity or failure [...]” (Wiley and Koene, 2016, p. 214)

This paper has two parts. In the first part (§2 and §3), I present a response to Wiley and Koene’s objection. I make the case that their attempt to establish a metaphysical equivalence between gradual and instant replacement uploading fails. The alleged equivalence hinges on the assumption that the rate of replacement of material parts is in general irrelevant for persistence. But this assumption is incompatible with a deeply entrenched view about the metaphysics of ordinary objects. It is almost universally accepted that ordinary material objects can survive *gradual* replacement of parts, but not *complete and sudden* replacement. I provide further justification for this view, and against Wiley and Koene’s objection, by showing that our judgments about the metaphysical significance of replacement rates are ultimately founded on differences in *immanent causation*.

In the second part of the paper (§4), I make a positive case for selective optimism about gradual uploading, relying on these differences in immanent causation. Gradual replacement uploading differs from instant and scan-and-copy uploading in involving a high, rather than a low, degree of immanent causation. It is furthermore plausible that this contrast in immanent causation plays a role for the persistence of human persons, just as it does for the persistence of ordinary material objects. Hence, the differences in immanent causation can provide the rational ground

² Olson (2017) dismisses selective optimism without much discussion: “I doubt whether anyone thinks that gradual uploading is metaphysically possible but all-at-once uploading is not” (Olson, 2017, p. 43).

for optimism about gradual replacement uploading, and pessimism about scan-and-copy uploading.

2. Wiley and Koene's Objection Against Selective Optimism About Uploading

Let's start by considering Wiley and Koene's objection against selective optimism in more detail. As alluded to above, Wiley and Koene use *instant replacement uploading* as an intermediary step linking gradual replacement and scan-and-copy uploading. Like gradual replacement uploading, instant replacement uploading involves replacing a brain's biological neurons with functionally equivalent silicon neurons. But in contrast to the gradual version, this happens from one moment to the next, rather than successively over several months, days, or hours. The following is a reconstruction of Wiley and Koene (2016)'s objection:

Wiley and Koene's Objection against Selective Optimism

1. There is no relevant metaphysical difference between scan-and-copy uploading and gradual replacement uploading.
2. There is no relevant metaphysical difference between instant replacement uploading and gradual replacement uploading.
3. By transitivity: there is no relevant metaphysical difference between scan-and-copy uploading and gradual replacement uploading.
4. If there is no relevant metaphysical difference between scan-and-copy uploading and gradual replacement uploading, then selective optimism about gradual replacement over scan-and-copy uploading is false.

∴ Selective optimism about gradual replacement uploading over scan-and-copy uploading is false.³

To simplify the discussion, we can follow Wiley and Koene in considering only hypothetical uploading cases where the uploading technology works flawlessly and also in focusing exclusively on cases of *destructive* replacement and scan-and-copy uploading, in which the relevant original biological hardware gets destroyed.

Prima facie, premise 1 does not seem plausible. In the case of instant replacement uploading, the end-product of the process is a siliconized brain (inside a human organism). In the case of scan-and-copy uploading, the outcome is the implementation of a simulation of a person's brain on a computer. Since the two results are rather different, the respective processes leading up to them will likewise differ in multiple ways. Why then, should we expect there to be no differences between the two procedures that could matter for personal identity?

Note, in response, that Wiley and Koene imagine a special variant of scan-and-copy uploading. In the envisioned scenario, the simulation is implemented on a

³ This argument likewise shows that Selective Optimism about gradual over instant replacement uploading is false, since premise 2 states that the two do not differ either.

neuromorphic computational architecture that is physically identical to the silicon brain which results from replacement uploading: “the brain is frozen, sectioned, and scanned, and the scan data are used to build an artificial brain composed of billions of prosthetic neurons [...]” (Wiley and Koene, 2016, p. 220). Wiley and Koene contend that, since both the process associated with scan-and-copy uploading, and the one associated with neuron replacement uploading, terminate in identical physical states, it is plausible that they are equivalent with regards to facts about personal identity. I will in the following focus on this specific version of scan-and-copy uploading.⁴

Is there not still a relevant difference between the two processes in that scan-and-copy uploading displaces the uploaded person potentially light-years away from the original, while instant replacement leaves them where they are? Wiley and Koene admit that there is a difference here. However, they hold that it is only one of degree. Even in replacement uploading, there is some spatial displacement between the biological and the siliconized brain, since the silicon-neurons and the original biological neurons do not share the exact same location. True, the spatial displacement occurs at the micron-level, rather than at the level of light-years. But still, the nature of the contrast is merely quantitative, concerning the magnitude of spatial displacement, rather than qualitative, concerning the contrast between spatial displacement vs. no spatial displacement. Next, Wiley and Koene make the case that it is implausible that a mere difference in the magnitude of spatial displacement could lead to a difference in whether the person undergoing uploading survives. Assuming otherwise requires either an arbitrary seeming spatial cut-off between adjacent cases, e.g. that personal identity is preserved only when the spatial displacement is, say, less than 93.264cm. Or alternatively, it requires an equally problematic smooth blend of partial identity failure/preservation as a function of spatial distance. Both options seem unpromising. They conclude that it is more reasonable to assume that spatial displacement is metaphysically irrelevant for personal identity, such that either any spatial displacement destroys the person or none does.

I will not discuss premise 1 further for now. My aim is instead to show that premise 2 is false, by demonstrating that there is in fact a metaphysically significant difference between *gradual* replacement uploading and *instant* replacement uploading (see §3). Wiley and Koene’s justification for premise 2 is the temporal analogue of the spatial case for premise 1. They imagine a spectrum of closely related replacement cases with ever faster rates. One may start with a replacement rate of, say, 1000 neurons per

⁴ Two comments: First, given that different types of uploading each allow for various manifestations, it may be better to formulate the above premises along the following lines: [1*] *There is no metaphysical difference between certain forms of scan-and-copy uploading and certain forms of instant replacement uploading*, and similarly for the other premises. We should then also formulate selective optimism about gradual uploading accordingly as the claim that it is possible to survive **certain forms of gradual uploading**, but not possible to survive **any form of scan-and-copy uploading**. I will in the following not explicitly mention this complication. Second, Wiley and Koene’s seem to assume here that persistence is an *internal* relation, i.e. one that is fully determined by the intrinsic character of the relata (Lewis, 1986). This is highly implausible. For instance, persistence is clearly subject to *causal* constraints. A future subject who is causally unrelated to me is not even a contender for being my successor, even if s/he is a perfect intrinsic duplicate of my future self.

second (according to Wiley and Koene's estimates, the gradual uploading procedure would then take around 1000 days). At the other end of the spectrum, we have instant replacement uploading where all neurons get exchanged from one moment to the next. Again, Wiley and Koene maintain that there cannot be a difference in identity preservation along this temporal spectrum, because assuming otherwise implies either a sharp cut-off or a smooth continuum ranging from full identity preservation to complete failure of identity preservation. And once more, they deem both options untenable. If they are right, there is no relevant metaphysical contrast between the cases along this spectrum that could explain a difference in survival. I am going to discuss this premise at length in the next section.

Let me finally give a short assessment of premises 3 and 4. As formulated, one might reject premise 3, since the expression *there is no relevant metaphysical difference between x and y* does not refer to a transitive relation. A chain of small, irrelevant differences may add up to a relevant difference. That notwithstanding, we can grant that if Wiley and Koene could establish the respective equivalences between instant replacement uploading and gradual replacement uploading or scan-and-copy uploading (i.e. premises 1 & 2), it would be difficult to deny the equivalence between gradual replacement uploading and scan-and-copy uploading (i.e. premise 3), independently of whether this premise is strictly entailed by the first two or not. So I will here not quibble about premise 3. Premise 4 should be uncontroversial. Selective optimism maintains that it is possible to survive gradual replacement uploading, but impossible to survive scan-and-copy uploading. Such facts about personal survival are not free-floating or primitive; they are grounded in some metaphysical difference between the two procedures. Therefore, if there is no such difference, then we either survive or fail to survive both, and selective optimism is false. Let us then have a closer look at the crucial premise 2.

3. Response to Wiley and Koene's Objection: Persistence and Immanent Causation

3.1 Material Replacement and the Persistence of Ordinary Objects

When assessing the plausibility of premise 2, the following observation should give us pause. Wiley and Koene's motivation for this premise is the claim that the rate at which the material parts of an object get replaced is in general irrelevant for the persistence of the object. This claim is entirely general and does not turn on the specifics of uploading or one's view of personal identity. If correct, it should apply equally to the ubiquitous exchange of material parts in ordinary objects, such as chairs, crystals, or cockatoos, all around us. However, our judgments about the persistence of such ordinary objects fly in the face of Wiley and Koene's verdict.

It is a platitude that material objects can persist through an exchange of material parts. One and the same tree loses its leaves in the winter and grows new leaves in the spring. When your car is being serviced, faulty old parts get exchanged with new ones. The car you pick up from the workshop is the original car with new parts, rather than

a new car with most of the old parts. It is furthermore generally assumed that it matters for an object's persistence whether material parts are being exchanged gradually or suddenly on a large-scale. This assumption is so firmly entrenched in the relevant debates in metaphysics that it is typically not even explicitly articulated. For instance, it is a central aspect of the most famous example of part replacement and persistence—the ship of Theseus—that the replacement happens gradually, over an extended period of time: “Consider the Ship of Theseus: a wooden ship that, *over the course of time, gradually* undergoes the replacement of each of its constituent planks.” (Rea, 1995, p. 532; my emphasis). Ordinary objects can survive *gradual* replacement of parts. The implicit flip-side of this claim is that such objects can *not* survive a process of *sudden and large-scale* replacement. If all cells of a tree were destroyed and replaced with new ones from one instant to the next, the original tree would have been replaced by a new tree. Equally, if the car mechanics discarded all of your car's old parts from one day to the next and reassembled the same model from brand new ones, you would be the lucky recipient of a new car.

That the difference between gradual and sudden large-scale material replacement is relevant for the persistence of ordinary objects is backed both by common sense and by the mainstream in metaphysics, and directly contradicts Wiley and Koene's assumption. We can turn this point into an explicit argument against premise 2 of Wiley and Koene's objection:

The Response from the Persistence of Ordinary Objects

1. Premise 2 of Wiley and Koene's objection presupposes that the rate of material replacement is irrelevant for the persistence of ordinary objects.
2. If the rate of material replacement is irrelevant for the persistence of material objects, then there is no difference in persistence between gradual and instantaneous material replacement for ordinary objects.
3. There is a difference in persistence between gradual and instantaneous material replacement for ordinary objects.

∴ Premise 2 is false and the rate of material replacement can be relevant for persistence.

In this argument, Premise 3 is where the action is at. To repeat the rationale for this premise: it is extremely plausible, and widely accepted by both experts and non-experts, that ordinary macro-objects can survive *gradual* replacement of material parts, but do not survive a process of sudden large-scale replacement.

It is natural to demand a deeper explanation for this contrast. Why, a skeptic may ask, should the rate of replacement matter for persistence? Indeed, imagine we could zoom in on the relevant object and observe the replacement process for different replacement rates at the micro-level. We would, it seems, observe the very same physical process unfolding at different speeds. But if the process is otherwise the same, why should a mere difference in speed make a difference for persistence (Chalmers, 2010)? The claim that the rate of replacement matters smacks of subjectivity and anthropomorphic bias (Wiley and Koene, 2016). Our intuitions about

metaphysical reality might simply be distorted by features of our human perspective, e.g. by the processing speed of our cognitive system.

These are legitimate concerns. However, they lose their bite once we realize that it is not differences in replacement rates *per se* which are the fundamental ground of differences in persistence. Rather, as I will aim to show in the next section, replacement rate is a surface indicator of a deeper phenomenon. The nature of the replacement process is correlated with the level of *immanent causation* connecting earlier stages of an object with potential successors of the object. And it is such differences in immanent causation which ultimately explain the relevant differences in persistence

3.2 Immanent Causation and Material Replacement

Immanent causation contrast with transeunt causation where one object causally affects another. In immanent causation, the causal links are located within the object itself (Zimmerman, 1996; Olson, 2016). What is the connection between immanent causation and persistence? The answer is that persisting objects immanently cause their own continued existence, and later states of a persisting object are typically immanently caused by its earlier states. Why does object *o* exist at t_2 ? Because *o* existed at t_1 . Why is object *o* red and round at t_2 ? Because *o* was red and round at the earlier time t_1 . As Eric Olson writes: "For a thing that exists now to exist in the future, then, it must cause itself to exist then, and the way it is now must to some extent cause it to be the way it is then. Or at least the existence and state of a thing in the future has to relate to its existence and state now by a chain of causal connections. And these connections must be immanent." (Olson, 2016, p. 56).

The significance of immanent causation for persistence also becomes apparent when we ask whether ordinary objects, such as a rock or a statue, could be brought back into existence after their destruction. This seems impossible (van Inwagen, 1978; Olson, 2016). Consider one of Michelangelo's statues whose parts have disintegrated and withered away over the centuries (compare also the examples in (van Inwagen, 1978) and (Olson, 2016)). Could we do something today that would bring Michelangelo's statue back into existence? Even if we somehow managed to collect up all the original matter and arranged it in the exact same way, we would not have thereby resurrected Michelangelo's original. Rather, we would have created a replica. And the same holds for other everyday objects. The reason why we cannot bring such objects back into existence after their destruction is that there is not a sufficient level of immanent causation between earlier states of the object and temporally later states of the reconstructed object. "If something tomorrow were exactly like your cat or your toothbrush is today, it wouldn't be your cat or your toothbrush unless there were a significant immanent causal connection." (Olson, 2016, p. 57)

A sufficiently high degree of immanent causation between earlier and later stages is a requirement for the persistence of ordinary material objects. To complete the case against premise 2 of Wiley and Koene's objection, we still need to connect immanent causation with the process of material replacement. The basic connection is straightforward, even though the details are somewhat complex. In a nutshell, gradual replacement is associated with high levels of immanent causation; sudden or

instantaneous replacement, on the other hand, is correlated with low or zero levels of immanent causation. Let us see why.

Distinguish an object's original or *old* parts from the replacing or *new* parts. We can say that the level of immanent causation between an object's state at one time and the later state of a potential successor is determined by the extent to which the successor state is caused by old parts compared to new parts. Roughly, if the later state is mostly caused by old parts, the level of immanent causation between earlier and later state is high. If it is mostly caused by new replacing parts, the level of immanent causation is low. (We are here ignoring other potential external forces that may affect an object's future states.) In the following, I will focus on cases of *large-scale* or *complete* material replacement, i.e. situations where most or all of an object's material parts are eventually exchanged. One might think that in cases of large-scale replacement, the level of immanent causation will inevitably approach a low level, independently of whether replacement is gradual or sudden. Since most or all old parts get exchanged in the end, it may seem as if in the final stages of both gradual and sudden replacement later states will be caused predominantly by new parts, since the majority of old parts are no longer present. However, this line of reasoning overlooks the fact that the replacing parts do not remain new parts forever. Over time, replacing parts get incorporated into the complex of parts that constitutes the object—new parts become old parts.

The interplay between the replacement of parts and their incorporation can lead to differences in immanent causation between slow and fast replacement in two different ways. First, incorporation of new parts requires that at each step a sufficient amount of the object's old parts do not get replaced. In other words, replacement needs to be *incremental*, in that each replacement step must be sufficiently *small-scale* (even when the replacement process as a whole is large-scale). Otherwise, immanent causation at intermediate steps is too low to support the object's persistence, and the residual old parts cannot integrate new parts into a complex of parts that still constitutes the original object. The incrementalist requirement directly separates gradual from *instantaneous* replacement. In instantaneous replacement, all old parts get replaced from one step to the next. Here, the degree of immanent causation is zero, since the relevant successor state is directly caused entirely by new parts. The incrementalist, stepwise small-scale replacement requirement already suffices to establish a significant metaphysical contrast between gradual and instantaneous replacement: gradual replacement can comply with the incrementalist requirement while instantaneous replacement cannot. In addition, a further difference between slow and rapid replacement can arise even for cases that comply with the incrementalist requirement. This leads to the second point.

Assume, as seems plausible, that incorporation of new parts takes a non-zero amount of time. When the replacement process is incremental and slow, replacing parts have enough time to become old parts. Since incorporation happens faster than replacement, the relevant causal contribution comes mostly from old parts, and the level of immanent causation is high throughout. However, when replacement happens faster than incorporation, subsequent stages are to a higher degree caused by new parts, and the level of immanent causation decreases. Therefore, as the gap

between rate of replacement and rate of incorporation between cases grows, there will be a correspondingly growing decrease in immanent causation.

To see this, consider the following toy model. Object o is composed out of 100 identical parts. Assume that each part of o is equally involved in causing future states of the object. Assume that the rate of incorporation of new parts is constant—new parts become old parts after, say, three units of time. We can calculate the level of immanent causation for different replacement rates, starting with a replacement rate of one old part per unit of time (identifying the level of immanent causation with the the number of old parts compared to the number of total parts that are involved in causing the relevant subsequent state). At t_{10} , after 10 units of time, 10 out of 100 old parts have been replaced by new parts. And seven new parts have become old parts; since it takes three units of time to become an old part, all new parts introduced at or before time t_7 have become old parts at t_{10} . That leaves 3 new parts and 97 old parts at t_{10} . So at this replacement rate, the ratio of old to total (i.e. old + new) parts involved at t_{10} in causing the subsequent state at t_{11} and the corresponding level of immanent causation is 0.97%. Consider next the level of immanent causation for a replacement rate of 5 parts per unit of time. At this rate, 50 old parts have been replaced at t_{10} . Again, all new parts introduced at or before step t_7 have now become incorporated. That means that there are 85 old and 15 new parts involved in causing the next state. Accordingly, the level of immanent causation for this replacement rate is 0.85%. At a replacement rate of 10 old parts per unit of time, the level of immanent causation drops to 0.7%. Etc. While reality is obviously more complex than the toy model, the model serves to illustrate the crucial point that an increase in replacement rate can lead to a decrease in immanent causation, even for incremental replacement processes.

To conclude, there are two ways for discrepancies in immanent causation between gradual and rapid replacement to arise. First, unlike instantaneous large-scale replacement, gradual replacement can comply with the incrementalist replacement requirement. Second, when replacement rate exceeds rate of incorporation, immanent causation decreases the faster the replacement occurs (assuming that the rate of incorporation is constant). These points show that there is a close connection between the nature of the replacement process and the degree of immanent causation, which in turn matters greatly for the persistence of ordinary objects. This undermines Wiley and Koene's case for the alleged metaphysical equivalence between gradual and instant replacement uploading on which their objection against selective optimism rests.

4. The Argument for Selective Optimism

The above provides a strong *negative* case against Wiley and Koene's objection. Can we utilize the same considerations to devise a *positive* case for selective optimism? The answer hinges on two additional questions. First, is there an analogous contrast in immanent causation between gradual replacement uploading and instant replacement or scan-and-copy uploading? Second, does this contrast matter for the persistence of *human persons*, just as it matters for the persistence of ordinary material objects? Both questions plausibly have positive answers. Yes, there is a parallel

contrast in immanent causation between gradual uploading and the two alternative forms of uploading. And yes, this contrast matters for the survival of the uploaded person. Selective optimism about gradual replacement uploading emerges as a natural view out of the considerations about immanent causation. For clarity, I distinguish two versions of selective optimism, one about gradual vs. instant replacement uploading, and one concerning gradual replacement uploading vs. scan-and-copy uploading.

Selective Optimism_{grad/instant}: It is possible to survive gradual replacement uploading, but not possible to survive instant replacement uploading.

Selective Optimism_{grad/s-&c}: It is possible to survive gradual replacement uploading, but not possible to survive scan-and-copy uploading.

Using these labels, we can formulate the case for selective optimism as follows. Gradual and instant replacement uploading differ in involving high vs. low levels of immanent causation. According to plausible views of personal identity, this makes a difference for the survival of the uploaded person. This yields Selective Optimism_{grad/instant}. Since scan-and-copy uploading also involves low levels of immanent causation, the same reasoning also gives us Selective Optimism_{grad/s-&c}.

4.1 Immanent causation and uploading

Consider first the claim that there is a difference in immanent causation between gradual and instant replacement uploading. In gradual replacement uploading, the biological neurons of a person's brain are gradually replaced with silicon neurons. Assume that the replacement process takes place over several weeks, such that only a small fraction of the brain's neurons gets exchanged at each stage. As the replacing silicon-neurons are actively involved in the function of the hybrid biological-silicon brain, they become old parts of the brain. Just like in the case of gradual replacement in ordinary objects, there is a consistently high level of immanent causation linking temporally adjacent states of the neural system throughout.

In the case of instant replacement uploading, on the other hand, the transition from biological to silicon brain happens in an instant. Here, the state of the silicon brain is not caused by earlier states of the biological brain. Instead, it is caused entirely by the new replacing silicon neurons. Since the silicon neurons have so far played no active role in the brain's functioning, they have not become incorporated into it and have not acquired the status of old parts. As a result, the final state of the biological brain and the initial state of the silicon brain are not linked by immanent causation.

This difference in immanent causation motivates the following diagnosis. Gradual replacement uploading involves one persisting piece of mental hardware. Since the replacement process is incremental and involves a sufficiently high degree of immanent causation between adjacent stages, there is one enduring material object which successively turns from a biological brain into a silicon brain.⁵ On the other

⁵ Some may balk at my use of terminology here, i.e. at labelling the continuous system containing increasing amounts of silicon neurons a 'brain', insisting that this term is reserved for a biological organ, not permitting of non-organic parts. More neutrally, we may instead refer to the system as the person's

hand, because instant replacement is non-incremental, there are two objects involved rather than one material continuant. As the later state of the silicon brain and the preceding state of the biological brain are not connected by immanent causation, the biological brain gets destroyed and is replaced by a functionally equivalent silicon brain. We can sum up the difference between instant and gradual replacement uploading as follows:

Brain Difference: Gradual uploading involves one piece of underlying mental hardware (an object that turns from an organic brain into an artificial brain); instant replacement involves two pieces of mental hardware (an organic brain and a separate artificial brain).

4.2 Personal identity and the brain

Our next question is whether the contrast in immanent causation and the resulting Brain Difference matters for the survival of the uploaded person. The answer depends to some extent on one's view of personal identity. We can distinguish between two broad approaches to the problem of personal identity: a biological approach and a mental approach. Both are best regarded as a *families* of views, held together by certain core theoretical commitments.

Biological theories start off from the truism that we are biological organisms, i.e. members of the species *homo sapiens* (Olson, 1997, 2007; Snowdon, 2014). This claim about our nature leads to a biological view of our persistence: we continue to exist as long as our organism is alive. Mental theories of personal identity, in contrast, are motivated by the idea that our persistence is essentially mental—we stretch as far as our mental life does, which in turn is intimately tied to the functioning of our brain (Lewis, 1976; Shoemaker, 1984; Parfit, 1984). It is common to distinguish between a *narrow* mental view and a *wide* mental view of personal identity (Parfit, 1984, 2008). Both require that later mental states of a person are *causally* related to their earlier mental states. The narrow view is more stringent about the nature of the causal relation; it demands that the source of the causal connection is a continuing mental hardware, i.e. our *brain*. The wide view allows any causal connection. If any account of personal identity deserves to be labelled 'the standard view', it is arguably the narrow mental view.

The biological account's attraction derives from its commonsensical answer to the question of what we are: human animals. The mental account's appeal lies in its intuitive account of personal survival. Suppose it was possible to surgically remove and destroy all parts of your organism while keeping your brain functioning normally. As your mental life would flow on, it seems hard to deny that you would continue to exist, although in a radically pared down form. So it seems that we can outlive our organism. Accordingly, Thomas Nagel writes: "[...] what is essential to my existence—is my functioning brain. [...] the brain is the only part of me whose destruction I could not possibly survive. The brain, but not the rest of the animal, is essential to the self." (Nagel, 1986, p. 40). To accommodate this compelling intuition

'mental hardware' or 'integrating apparatus' (Peacocke, 2014, §III). Ultimately, this is a merely verbal issue. For convenience, I will continue to refer to the system as a 'brain'.

about personal survival, some proponents of the biological view hold that not just you, but your organism survives as long as your brain does, since the brain is the organ that sustains your kind-specific functions and capacities (Madden, 2016).

4.3 The Argument for Selective Optimism

Both according to the influential narrow mental view of personal identity and according to certain versions of the biological view, we survive as long as our brain exists and is able to support our mental life. Combined with the outcome of §4.1, i.e. Brain Difference, this suggests that we can survive gradual replacement uploading but not instant replacement uploading. Gradual uploading involves one persisting piece of mental hardware—an entity that slowly turns from a biological brain into a silicon brain. Since the replacing silicon neurons are by assumption functionally equivalent to the biological neurons, the hybrid brain will support a continuous mental life throughout, according to an orthodox view in the philosophy of mind (see e.g. Chalmers, 2011, 2012)).⁶ Similarly, the hybrid brain continuously supports the relevant kind-specific functions on which our survival depends according to a certain version of the biological account (Madden, 2016). In contrast, in the case of instant replacement uploading, your brain gets destroyed and is replaced by a numerically distinct silicon brain. But if, as the above views of personal identity maintain, your survival depends on the continuity of your mental hardware, then you do not survive this procedure.

Taken together, these points give us the following argument for Selective Optimism_{grad/instant}:

The Argument for Selective Optimism_{grad/instant}

1. Gradual replacement uploading involves high levels of immanent causation; instant replacement uploading involves low levels of immanent causation.
2. The differences in immanent causation entail that there is continuity of mental hardware in gradual replacement uploading but not in instant replacement uploading.
3. Continuity of mental hardware and its functions is the necessary condition for the survival of human persons.
4. Gradual replacement uploading satisfies the necessary condition for our survival; instant replacement uploading does not satisfy the necessary condition for our survival.
5. If (4) is true, then it is possible to survive gradual replacement uploading, but not possible to survive instant replacement uploading.

∴ It is possible to survive gradual replacement uploading, but not possible to survive instant replacement uploading.

⁶ Although widely accepted, this view of the mind is not uncontroversial. Biological accounts, for instance, imply that there will likely not be a seamless continuity of mental life in gradual uploading (Block, 1978, 2009; Searle, 1992). It is beyond the scope of this paper to settle this dispute.

The two preceding sections contain the motivation for the core premises in this argument. §4.1 made the case that there is a difference in immanent causation between gradual and instant replacement uploading, which results in a contrast in the continuity of the brain. §4.2 suggested that this contrast matters for the survival of the uploaded person.⁷

If a lack of immanent causation results in a failure of personal survival in instant replacement uploading, the same holds for scan-and-copy uploading, since here the level of immanent causation connecting the biological brain and the artificial brain is equally low. Just like in instant replacement uploading, the later state of the artificial brain is not immanently caused by the earlier state of the biological brain. Hence, Brain Difference also holds for scan-and-copy uploading. (The main difference between scan-and-copy and instant replacement uploading is that the former process involves a (larger) spatial and a temporal gap between the final stage of the biological brain and the initial stage of the artificial brain.) And as before, Brain Difference suggests that the person does not survive scan-and-copy uploading. This leads to the following argument for Selective Optimism_{grad/s-&c}:

The Argument for Selective Optimism_{grad/s-&c}

1. Gradual replacement uploading involves high levels of immanent causation; scan-and-copy uploading involves low levels of immanent causation.
 2. The differences in immanent causation entail that there is continuity of mental hardware in gradual replacement uploading but not in scan-and-copy uploading.
 3. Continuity of mental hardware and its functions is the necessary condition for the survival of human persons.
 4. Gradual replacement uploading satisfies the necessary condition for our survival; scan-and-copy uploading does not satisfy the necessary condition for our survival.
 5. If (4) is true, then it is possible to survive gradual replacement uploading, but not possible to survive scan-and-copy uploading.
-
- ∴ It is possible to survive gradual replacement uploading, but not possible to survive scan-and-copy uploading.

Relying on the connection between immanent causation and persistence, I have proposed a positive argument for selective optimism about gradual replacement uploading. The differences in immanent causation between gradual replacement and instant replacement uploading or scan-and-copy uploading suggest that we can survive the former but not the latter two. This case for selective optimism is not incontrovertible. My argument relied on certain assumptions about personal

⁷ For alternative arguments against scan-and-copy uploading see e.g. (Corabi and Schneider, 2014; Goldwater, 2021; Olson, 2022). Goldwater (2021) contends that gradual uploading should not be counted as a genuine form of uploading at all. But this seems like a verbal issue. Gradual replacement uploading conforms to the initial definition of uploading in §1, and it promises to deliver many of the benefits that motivate uploading optimists, such as potential life-extension. So our use of terminology seems legitimate.

identity and the mind. While plausible and widely shared (Bourget and Chalmers, 2023), these assumptions are actively debated.

It should be noted, however, that the standard biological account of personal identity also predicts that there can be a difference in personal identity between gradual replacement uploading and scan-and-copy uploading and may therefore lead to selective optimism. On the standard biological view, the brain is not essential to the survival of the organism. Hence, one can survive destructive gradual replacement uploading, which only affects the biological brain and not necessarily the rest of the organism, but not destructive scan-and-copy uploading, which, we may assume, destroys the entire organism.

My aim here has not been to make a definitive case that gradual replacement uploading is survivable while the other two forms are not. Rather, I wanted to show that it is reasonable to think that there is a significant metaphysical difference between the different forms of uploading, and that this difference may matter for our survival. If given the choice between various forms of uploading, I would rather choose the gradual version, and I do not believe that this preference is irrational.

5. Selective Optimism about Uploading and Vagueness

Endorsing selective optimism about gradual uploading returns us to the dilemma mentioned in Wiley and Koene's objection. When considering a continuous spectrum of ever decreasing levels of immanent causation, we are committed either to a sharp cut-off in identity preservation somewhere along this spectrum, or to a range of cases of partial or indeterminate survival.

We can agree with Wiley and Koene that these options do not look particularly attractive. However, having to accept one of them may simply be where we are. If ordinary objects do indeed survive gradual but not sudden large-scale material replacement, then we are already committed to one of these alternatives for these objects, independently of the discussion around uploading. Wiley and Koene motivate their position by Occamist parsimony considerations—they contend that treating the different forms of uploading equally is theoretically more parsimonious than the alternative. But if our judgments concerning ordinary objects are to be trusted, then their assumption that the contrast between gradual and instant replacement makes no difference to our survival is in fact the costly option. We need to provide extra justification for the claim that there is some underlying metaphysical difference between ourselves and ordinary objects, such that we deserve special theoretical treatment.

The above dilemma is familiar from discussions of vagueness. In a typical case of vagueness, it seems implausible that there are sharp boundaries in nature, separating cases that differ physically in only trivial ways. A single grain of sand does not seem to demarcate the border between heaps and non-heaps. That notwithstanding, there are sophisticated defences of such precise joints of nature. Epistemicists about vagueness hold that there are sharp but unknowable cut-offs (e.g. Williamson, 1994). Similarly, supervenientists accept sharp boundaries for

precifications of our ordinary concepts (e.g. Fine, 1975; Lewis, 1986). Alternatively, there may be a range of borderline cases where the relevant concept neither clearly applies nor clearly fails to apply. This option does not in itself commit us to the claim that any less than maximal level of immanent causation leads to a partial failure of identity preservation. It seems more plausible that there may be threshold level of immanent causation, such that above this threshold identity is fully preserved. Only below the threshold do we encounter borderline cases of survival. The phenomenon of higher-order vagueness suggests that the exact location of this threshold is itself indeterminate.

Admittedly, indeterminate persistence is much harder to accept in our own case, than in the case of ordinary material objects. When considering our own survival, there is a strong intuition that we will either definitely survive a certain vicissitude or that we definitely won't—*tertium non datur*. In other work, I have argued that this intuition has a psychological explanation, tracing back to the special character of first-person imagination (Weber, 2024a). We typically imagine hypothetical cases involving our own survival from the inside. From this perspective, indeterminate survival is indeed inconceivable, since it would require imagining having a borderline case of experience. But the special first-person inconceivability of indeterminate survival does not entail its metaphysical impossibility (Weber, 2024a,b). So there may indeed be cases of indeterminate personal survival. If we look closely enough, we can see that most natural domains are subject to vagueness. It would be surprising if things were different in our own case, since our survival is simply another natural phenomenon.

Acknowledgements

For helpful comments and discussion, I would like to thank Jesse Hambly, Phil Pickering, and Wolfgang Schwarz. This research was supported by a fellowship from the Australian Research Council (DE220101158).

References

- Block, N. (1978). Troubles with functionalism. *Minnesota Studies in the Philosophy of Science* 9, pp. 261–325.
- Block, N. (2009). Comparing the major theories of consciousness. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences*, Volume 4, Cambridge, MA: Blackwell.
- Bourget, D. and D. Chalmers (2023). Philosophers on philosophy: The 2020 philpapers survey. *Philosophers' Imprint* 23(1).
- Chalmers, D. (2010). The singularity: a philosophical analysis. *Journal of Consciousness Studies* 17, 7–65.
- Chalmers, D. (2011). A computational foundation for the study of cognition. *Journal of Cognitive Science* 12(4), pp. 325–359.
- Chalmers, D. (2012). The varieties of computation: A reply. *Journal of Cognitive Science* 3, pp. 211–248.
- Corabi, J. and S. Schneider (2014). If you upload, will you survive? In R. Blackford and D. Broderick (Eds.), *Intelligence Unbound*, Wiley.

- Goldwater, J. (2021). Uploads, faxes, and you: Can personal identity be transmitted? *American Philosophical Quarterly* 58(3), pp. 233–250.
- Fine, K. (1975). Vagueness, truth and logic. *Synthese* 30(3-4), pp. 265–300.
- Lewis, D. (1976). Survival and identity. In A. O. Rorty (Ed.), *The Identities of Persons*, Berkeley: University of California Press.
- Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Basil Blackwell.
- Madden, R. (2016). Human persistence. *Philosopher's Imprint* 16(17), pp. 1–18.
- Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.
- Olson, E. T. (1997). *The Human Animal: Personal Identity without Psychology*. Oxford: Oxford University Press.
- Olson, E. T. (2007). *What Are We? A Study in Personal Ontology*. New York: Oxford University Press.
- Olson, E. (2016). Immanent causation and life after death. In G. Gasser (Ed.), *Personal Identity and Resurrection*, New York: Routledge.
- Olson, E. T. (2017). The central dogma of transhumanism. In B. Berčić (Ed.), *Perspectives on the Self*, Faculty of Humanities and Social Sciences University of Rijeka.
- Olson, E. T. (2022). The metaphysics of transhumanism. In K. Hübner (Ed.), *Human: A History*, Oxford, Oxford University Press.
- Parfit, D. (1971). Personal identity. *Philosophical Review* 80, pp. 3–27.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon.
- Parfit, D. (2008). Persons, bodies, and human beings. In T. Sider, J. Hawthorne, and D. W. Zimmerman (Eds.), *Contemporary Debates in Metaphysics*, Cambridge, MA: Blackwell.
- Peacocke, C. (2014). *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*. Oxford: Oxford University Press.
- Rea, M. (1995). The problem of material constitution. *Philosophical Review* 104, pp. 525–52.
- Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge: MIT Press.
- Shoemaker, S. (1984). Personal identity: A materialist's account. In R. Swinburne and S. Shoemaker (Eds.), *Personal Identity*, Oxford: Blackwell.
- Snowdon, P. F. (2014). *Persons, animals, ourselves*. Oxford: Oxford University Press.
- van Inwagen, P. (1978). The possibility of resurrection. *International Journal for Philosophy of Religion* 9(2), pp. 114–121.
- Weber, C. (2024a). *De Se Modal Illusions*. Unpublished manuscript.
- Weber, C. (2024b). Determinacy of the Self and Imagination from the Inside. Unpublished manuscript.
- Wiley, K. B. (2014). *A Taxonomy and Metaphysics of Mind-Uploading*. Los Angeles, CA: Humanity+ Press and Alautun Press.
- Wiley, K. B. and R. A. Koene (2016). The fallacy of favouring gradual replacement mind uploading over scan-and-copy. *Journal of Consciousness Studies* 23(3–4), pp. 215–235.
- Williamson, T. (1994). *Vagueness*. London: Routledge.
- Zimmerman, D. W. (1996). Immanent causation. In J. Tomberlin (Ed.), *Philosophical Perspectives 11: Mind, Causation, and World: Metaphysics*, Cambridge, MA: Blackwell.