

The Personal Identity Dilemma for Transhumanism

RALPH STEFAN WEIR

Abstract

Transhumanists claim that futuristic technologies will permit you to live indefinitely as a nonbiological ‘posthuman’ with a radically improved quality of life. Philosophers have pointed out that whether some radically enhanced posthuman is *really you* depends on perplexing issues about the nature of personal identity. In this paper, I present an especially pressing version of the personal-identity challenge to transhumanism, based on the ideas of Derek Parfit. Parfit distinguishes two main views of personal identity, an intuitive, nonreductive view and a revisionary, reductive view. I argue that the standard rationale for *wanting* to become a posthuman makes sense only if the intuitive view is correct, but that the standard rationale for thinking that it is *possible* to become a posthuman makes sense only if the revisionary view is correct. Following this, I explain why the obvious responses are unsatisfactory or imply the need to rethink transhumanism in ways that make it much less radical and less appealing.

1. Introduction

Humans face many familiar limitations: physical and cognitive shortcomings, aging, disease, and death within a matter of decades. Transhumanists argue that rapid technological progress presents us with an opportunity to overcome these limitations, living vastly extended lives as radically enhanced ‘posthumans’ with a profoundly improved quality of life.¹ The first steps towards the posthuman future are expected to involve comparatively moderate enhancements such as brain implants that increase intelligence or nanomedicines that extend life. But according to many leading transhumanists,

¹ Following the *Transhumanist FAQ* I use ‘posthuman’ for beings who have undergone such radical enhancements that they are ‘no longer unambiguously human by our current standards’, and ‘transhuman’ for ‘an intermediary form between the human and the posthuman’ (Humanity+, 2024).

doi:10.1017/S0031819124000068 © The Author(s), 2024. Published by Cambridge University Press on behalf of The Royal Institute of Philosophy. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

later stages will involve much more radical enhancements, requiring the transference of one's mind to a computer ('uploading') or the gradual replacement of one's nervous system with artificial components ('neuron replacement therapy' or 'neural prosthesis').

Transhumanists present two motivations for replacing the nervous system with artificial components. First, there is longevity. As Nick Bostrom (2014, p. 60) observes, our brains 'start to permanently decay after a few decades of subjective time' whereas 'microprocessors are not subject to these limitations.' Transhumanists draw the conclusion that, in Randal Koene's words 'ultimately, it is our biology, our brain, that is mortal' (Piore, 2014). Furthermore, many join Ray Kurzweil (2005, p. 567) in hoping that 'a nonbiological existence' will provide 'the means of "backing ourselves up" ... thereby eliminating most causes of death as we know it.'

Secondly, there is perfectibility. According to Bostrom (2014, p. 60) 'the potential for intelligence in a machine substrate is vastly greater than in a biological substrate.' Hence, Kurzweil (2005, p. 344) predicts that 'although we are likely to retain the biological portion [of our intelligence] for a period of time, it will become of increasingly little consequence.' Likewise, Elise Bohan (2022, p. 42) states that inorganic intelligence is not only 'the most resilient in the long term' but also that it 'will evolve much further' than organic intelligence.

Not all who identify as transhumanists consider the wholesale replacement of the nervous system by artificial components essential to their project. The *Transhumanist FAQ* states that:

Posthumans could be completely synthetic artificial intelligences, or they could be enhanced uploads ... or they could be the result of making many smaller but cumulatively profound augmentations to a biological human. (Humanity+, 2024)

The third option leaves room for a posthuman future in which we retain augmented biological human bodies. Nonetheless, the replacement of the nervous system with technology is sufficiently central to leading versions of transhumanism to merit careful evaluation. It is this radical form of transhumanism that I focus on here. The degree to which the argument presented extends to more moderate versions of transhumanism will depend on details that I set aside here.

Three questions arise (cf. Chalmers, 2014; Olsen, 2017). Is it technologically possible to reproduce the functions of one's nervous system using artificial components? If this *is* possible, will the result be a conscious person, rather than an unconscious automaton? And if the result *is* a conscious person, will it be *oneself*?

The Personal Identity Dilemma for Transhumanism

The promise of transhumanism, as it is understood here, requires a positive answer to each question. This paper is about question three: assuming that radical enhancements are technologically possible, and that the result is a conscious person, will that person be *oneself*?

Philosophers have pointed out that the answer depends on controversial issues about personal identity.² In what follows, I present an especially pressing version of the personal-identity challenge to transhumanism, based on the ideas of Derek Parfit.³ Parfit distinguishes two main views of personal identity, an intuitive, non-reductive view on which personal identity is a ‘further fact’ over and above our psychological and physical characteristics; and a revisionary, reductive view on which personal identity is just a matter of certain psychological and/or physical patterns.

I argue that the standard rationale for *wanting* to become a posthuman makes sense only if the intuitive view is correct. For as Parfit argues, it is only on the intuitive view that personal survival into the future matters. On the other hand, I argue that the standard rationale for thinking that it is *possible* to become a posthuman makes sense only if the revisionary view is correct. For transhumanists defend the viability of technologies such as uploading or neural prosthesis by appeal to the idea that personal identity is just a matter of psychological and/or physical patterns.

For this reason, transhumanists face a dilemma. They can: i) adopt the revisionary view of personal identity and explain why we should want to survive as posthumans in the first place; or ii) adopt the intuitive view and explain why we should think it possible to survive as posthumans. If neither of these options works, transhumanists might explore a third option: iii) abandon the standard rationale for transhumanism in favour of one that does not involve *our own* survival.

Transhumanists who adopt option i) might appeal to existing strategies for resisting Parfit’s claim that on the revisionary view of personal identity, survival does not matter (e.g., Lewis, 1983; Johnston, 2003; Sosa, 2003). In response, I argue that the main strategies for resisting Parfit’s claim are either defective or are incompatible with transhumanism. Transhumanists who adopt option ii) might appeal

² In addition to Chalmers (2014) and Olsen (2017), see Corabi and Schneider (2012), Liz Stillwaggon Swan and Joshua Howard (2021), Pigliucci (2014), Walker (2014), Weir (2018, 2024), Schneider (2019), and Goldwater (2021).

³ I outline this challenge more briefly and compare it to other influential challenges to transhumanism in Weir (2024).

to David Chalmers' (2014) argument that uploading will work irrespective of our theory of personal identity. But I give reasons for thinking that Chalmers' argument fails. Transhumanists who adopt option iii) must come up with an alternative rationale for turning ourselves into posthumans. I argue that such a rationale is likely to have considerably narrower appeal, and to favour significantly less radical enhancements than transhumanism as it is typically understood.

2. Parfit on Personal Identity

Parfit's is by far the most influential work on personal identity in recent history. It is even cited in the *Transhumanist FAQ* as the leading work on the topic (Humanity+, 2024). So transhumanists ought to be interested in the consequences of Parfit's arguments for transhumanism. Parfit's arguments concern a distinction between two main views of personal identity (see Parfit, 1971, fn. 11; 1976, p. 227; 1984, p. 217). There is an intuitive view, supported by philosophers such as Joseph Butler (1736) and Richard Swinburne (1973–4), according to which personal identity is an irreducible, all-or-nothing matter. There is also a revisionary view, defended by philosophers such as William Hazlitt (1805) and by Parfit himself, according to which personal identity is reducible to degrees of psychological and/or physical continuity.

To get a sense of the intuitive view of personal identity, consider the following scenario. You are about to undergo a 'hemisphere upgrade', a procedure that will replace half your brain with powerful silicon circuits. You are confident that *someone* will wake up with half your brain, and many of your psychological traits, after the procedure. But you naturally ask yourself, 'Will that person be *me*?' Intuitively, this question must have a definite answer one way or the other. That is, it is very natural to think that either *you* will have the experience of waking up with the enhanced brain, or *someone else* will. There does not seem to be any room for a partial or indeterminate result.

The intuitive view of the identity of persons, so described, contrasts with the intuitive view of the identity of many other kinds of entity, such as nations or ships (to give two classic examples). It is plausible, as Parfit (1982, p. 228) observes, that 'the survival of a nation just consists in various kinds of continuities: demographic, territorial, cultural, political', and that when the continuities 'hold to intermediate degrees, there may be no answer to the question, "Does the same nation still exist?"' For example, it is plausible that

The Personal Identity Dilemma for Transhumanism

there is no definite answer to the question ‘Is post-1066 England the same nation as pre-1066 England?’ For this depends on how much continuity we demand before we count something as the ‘same nation’, and our decision is bound to be partly arbitrary.

According to the revisionary view of personal identity, defended by philosophers such as Hazlitt and Parfit, the identity of persons is in fact just like the identity of nations and ships. That is, the survival of a person just consists in various kinds of continuities: cognitive, emotional, physiological; and when the continuities hold to intermediate degrees, there may be no answer to the question ‘does the same person still exist?’ For example, the answer to the question ‘Will the person who wakes up with half your brain be you?’ depends on how much continuity we demand before we count someone as the ‘same person’, and our decision is bound to be partly arbitrary.

If the revisionary view of personal identity is correct then once we have all the information about the relevant continuities, the question ‘does the same person still exist?’ is ‘in the belittling sense, merely verbal’ because it depends on our decision about how much continuity we demand before we count someone as the ‘same person’ (Parfit, 2001, p. 25). This is one respect in which the revisionary view is counterintuitive. A second respect in which the revisionary view is counterintuitive is that it seems to make personal identity depend on extrinsic factors about whether the continuities have taken a branching form.

To see this, compare two scenarios. In the ‘single-upload’ scenario, your psychological states are transferred onto Computer 1, in a way that preserves the continuities relevant to personal identity.⁴ If the revisionary view of personal identity is correct, the upload on Computer 1 is therefore you. (We can assume, for the sake of the example, that it really is possible to recreate a person’s psychological states on a computer, though of course this is far from obvious.) In the ‘double-upload’ scenario, your psychological states are transferred onto Computer 1 *and* Computer 2: the continuities have taken a branching form. What happens to *you* in the double-upload scenario? It seems impossible that the uploads on Computers 1 and 2 are *both* you. For they are not identical to one another, and identity

⁴ Some proponents of the revisionary view think that personal identity requires greater physical continuity than uploading permits. But the point made in this paragraph can also be made using cases that involve a high degree of physical continuity such as the double-brain-transplant scenario discussed by Parfit (1984, pp. 254–64; 2001, pp. 39–49). I use the example of uploading because of its relevance to transhumanism.

is a transitive relation. It also seems impossible that just one upload is you. For neither has a greater claim to be you than the other. So it seems that *neither* upload in the double-upload scenario is you. Curiously then, on the revisionary view of personal identity, if your psychological states are transferred to Computer 1 only, the upload on Computer 1 will be you, but if your psychological states are transferred onto Computer 2 as well, neither the upload on Computer 1 nor the upload on Computer 2 will be you. In other words, if the revisionary view is correct, then personal identity depends on extrinsic factors about whether the continuities have taken a branching form.

Parfit argues that we should accept the revisionary view of personal identity in spite of its counterintuitive consequences. Parfit's main reason for favouring the revisionary view is that the intuitive view seems to require something like an indivisible soul to determine which person is you in situations where the continuities hold to intermediate degrees or branching has occurred. Butler and Swinburne accept that their position requires an indivisible soul but Parfit believes that no such thing exists.⁵

According to Parfit's (1984, pp. 206–7) favoured version of the revisionary view, personal identity is non-branching psychological continuity. A future person is 'psychologically continuous' with you if and only if their psychological states are joined to yours by chains of similarity and causal dependence. The 'non-branching' clause ensures that in cases like the double-upload scenario, multiple persons do not simultaneously satisfy the criteria for being 'you'. This is one formulation of the 'psychological continuity theory', the most popular version of the revisionary view of personal identity. The main alternatives to the psychological continuity theory say that personal identity consists in some kind of physical continuity.

Parfit argues that the revisionary view of personal identity has an important practical consequence. If personal identity is reducible to psychological and/or physical continuities then, Parfit argues, personal identity *does not matter*. That is, if the revisionary view is correct, then you should not place any value on your future existence or on the future existence of any other person.

⁵ An alternative version of the intuitive view says that personal identity involves a 'further fact' over and above the other facts, but not a further *thing* such as a soul. Everything I say about indivisible souls in what follows applies equally to irreducible personal-identity facts (cf. Parfit, 1984, p. 210).

The Personal Identity Dilemma for Transhumanism

If Parfit is right, then the revisionary view is even more counter-intuitive than it seemed to begin with. For it is very natural to attribute a high degree of value to one's future existence, as well as to the future existence of other persons. I do not just hope that someone with my psychological and physical characteristics will wake up tomorrow, I hope that *I* will do so. I do not just want persons with the psychological and physical characteristics of my friends and family to be happy, I want *those very persons* to be happy. The first- and second-personal attitudes at the centre of most persons' psychological lives suggest that personal identity matters very much indeed. For this reason, one might reasonably suspect that any theory on which personal identity does not matter must have gone wrong somewhere. Nonetheless, Parfit's *conditional* claim – that *if* the revisionary view of personal identity is correct, *then* personal identity does not matter – is hard to resist.

One reason for thinking that on the revisionary view, personal identity does not matter, concerns cases where the continuities hold to intermediate degrees like the hemisphere-upgrade scenario. If personal identity consists in psychological continuity, then it is up to us to decide whether the degree of continuity between you and the person who wakes up in this scenario is sufficient for us to count that person as 'you'. Because our decision will be partly arbitrary, it seems that nothing of practical importance could depend on it (Parfit, 1984, p. 241).⁶

A second reason for thinking that, on the revisionary view, personal identity does not matter, concerns cases of branching like the double-upload scenario.⁷ If personal identity consists in psychological continuity, it seems that you would survive having your psychological states uploaded onto Computer 1 only, but you would not survive having your psychological states uploaded onto both Computer 1 and Computer 2. And so, if personal identity matters, you should be happy to undergo single-uploading, but strongly opposed to double-uploading. But this, Parfit argues, would be absurd. If having your psychological states uploaded onto Computer 1 is as good as ordinary survival, the addition of the upload on Computer 2 cannot undermine this.

⁶ A similar point can be made with respect to cases where the continuities are caused in an unusual way, as in Parfit's teletransportation scenario, discussed at the same location.

⁷ Again, the same point can be made with respect to double-brain-transplant scenarios discussed by Parfit (1984, pp. 254–64; 2001, pp. 39–49).

It might seem outrageous to suggest that personal identity does not matter. But Parfit would say that this is only because we usually think of personal identity as a deep, all-or-nothing matter, of the sort that would require something like an indivisible soul. If the revisionary view is correct, then the sense in which some future person can be 'you' is much shallower than we usually imagine. So it should not surprise us that, if the revisionary view is correct, your future existence does not have the enormous value that you ordinarily attribute to it.

Parfit (1982, p. 229) thinks that it makes sense to place *some* value on the existence of the psychological continuities in which, in non-branching cases, one's future existence consists. For example, you can rationally want some future person to have memories of your experiences, intentions to carry out your plans, and so on. But you should be indifferent about whether that person is you. Likewise, you can rationally wish for the health and happiness of persons psychologically continuous with your friends and family. But you should be indifferent about whether those persons are your actual friends and family. Parfit adds that the degree of value that it makes sense to place on the continuities will be modest compared to the value that we ordinarily place on personal identity, and will diminish as the continuities grow weaker. So although on Parfit's view one can have some concern about which persons will exist in the future, that concern ought to be lower in degree and different in kind to the sort of concern that would make sense if the intuitive view of personal identity were correct.

3. The Rationale for Transhumanism

Susan Schneider (2019, pp. 72–73) maps out our future trajectory, according to influential versions of transhumanism, as follows. To begin with, we are unenhanced natural humans. Following this, we will undergo moderate enhancements, such as brain implants that increase intelligence or nanomedicines that extend life. We are still humans, but we have begun a process of transformation into something else: we are 'transhumans'. In the next stage, the enhancements become so radical that we are no longer 'unambiguously human by our current standards' (Humanity+, 2024). For example, our minds have been uploaded onto computers or our nervous systems replaced by silicon circuits. We are now nonbiological 'posthumans' with indefinite lifespans. Finally, further enhancements equip us with the cognitive abilities of superintelligent AI. We are not just

The Personal Identity Dilemma for Transhumanism

posthumans, but superintelligent posthumans, beings whose nature and experience we can, at present, hardly imagine.

Transhumanists represent this transformation as highly desirable. For example, in his 'Letter from Utopia', Bostrom (2008, p. 7) advertises the posthuman future as one in which 'every second is so good that it would blow your mind had its amperage not first been increased.' In a similar vein, Kurzweil promises that 'We're going to be funnier. We're going to be sexier... We're going to expand the brain's neocortex and become more godlike' (Kurzweil and Miles, 2015, pp. 24–5). And Bohan (2022, p. 250) expresses the belief that we're heading to 'a world in which our minds and bodies are digital, our experiences are virtual, and reality is much more of a choose your own adventure game.'

Transhumanists also represent the ordinary alternative – death in decades or less – as very undesirable. A recurrent motif in transhumanist literature is that death is a great tragedy, including when it occurs naturally in old age (see e.g., de Grey, 2007). For this reason many transhumanists see the development of life-extending enhancements as a matter of 'moral urgency'. Hence, Bostrom says:

150,000 human beings on our planet die every day, without having had any access to the anticipated enhancement technologies that will make it possible to become posthuman. The sooner this technology develops, the fewer people will have died without access... transhumanism stresses the moral urgency of saving lives. (Bostrom, 2005, pp. 11–13)

Similarly Kurzweil (2005, pp. 650–1) claims that 'we have the means right now to live long enough to live forever' but laments the fact that 'most baby boomers won't make it because they are unaware of the accelerating aging processes in their bodies and the opportunity to intervene.'

These examples reflect what I call the 'standard' rationale for transhumanism: we should want to become posthumans because doing so will allow us to defer death and to enjoy a vastly improved quality of life. This is not the only rationale that has been given for developing radical enhancements such as neural prosthesis or uploading, but it is shared by many leading transhumanists and has no comparably influential competitor.

But if becoming a posthuman by means of enhancements such as neural prosthesis or uploading is desirable, it is not obvious that it is possible. For one thing, it is not obvious that it is possible to reproduce the functions of one's nervous system with artificial components such as silicon chips. If some neurological processes function in a

noncomputable way, for example, then a computational simulation of one's nervous system will be impossible. Furthermore, if it is possible to reproduce the functions of one's nervous system with artificial components, it is not obvious that the result will be a conscious person. This depends, amongst other things, on whether conscious experience requires a biological substratum. Most importantly, for the purposes of this paper, even it is possible to reproduce the functions of one's nervous system using artificial components, and if the result is a conscious person, is not obvious that that person will be oneself.

In addition to a rationale for wanting to become a posthuman, therefore, transhumanists must provide a rationale for thinking that it is possible to do so. This means, among other things, explaining why one should expect the person who wakes up after neural prosthesis or uploading to be oneself.

In support of the thesis that the person who wakes up after neural prosthesis or uploading will be oneself, transhumanists tend to appeal to the revisionary view of personal identity (see e.g., Kurzweil, 1999, p. 383; 2005, pp. 675–8; Schneider, 2019, pp. 89–143; Humanity+, 2024). For example, in discussing the possibility of uploading, the *Transhumanist FAQ* says:

A widely accepted position is that you survive so long as certain information patterns are conserved, such as your memories, values, attitudes, and emotional dispositions, and so long as there is causal continuity so that earlier stages of yourself help determine later stages. (Humanity+, 2024)

The position described here is the psychological-continuity theory of personal identity. Kurzweil (2005, p. 678) also invokes the revisionary view when he appeals to the idea that 'I am principally a pattern that persists in time.' The suggestion is that if personal identity consists in certain psychological and/or physical patterns, and radical enhancements preserve those patterns, then radical enhancements preserve personal identity. Likewise, in his recent contribution to the literature on uploading, Watanabe Masataka (2022, p. 151) echoes Locke's memory-based psychological-continuity theory saying that 'we consider ourselves to be us because we retain memories' and that 'if we were to wake up in a machine that retained all of our memories ... we would not pause and wonder who we are.'

In summary, we should want to become posthumans because doing so will allow us to defer death and go on living in the agreeable conditions of the posthuman world. And we should think that

The Personal Identity Dilemma for Transhumanism

it is possible to become posthumans – or at any rate, that the nature of personal identity poses no obstacle to our becoming posthumans – because there is nothing more to personal identity than psychological and/or physical continuities of the sort that enhancements such as neural prosthesis and uploading are supposed to preserve.

4. The Dilemma

Most people find the idea of living on in an improved state, rather than dying, attractive. So the rationale for wanting to become a posthuman is at least *prima facie* reasonable.⁸ At the same time, the revisionary view of personal identity, and the psychological continuity theory in particular, are popular. To that extent, the rationale for thinking that it is possible to become a posthuman is *prima facie* reasonable as well. It is no surprise, therefore, that many people are persuaded that transhumanism is a good idea.

This optimistic picture is, however, misleading. For Parfit's arguments suggest that the rationale for wanting to become a posthuman, and the rationale for thinking this possible, rest on opposing views of personal identity.

The rationale for wanting to become a posthuman presupposes that one has a significant interest in one's future existence. So if Parfit is right, the rationale for wanting to become a posthuman presupposes the intuitive, non-reductive view of personal identity. By contrast, the rationale for thinking that it is possible to become a posthuman presupposes that personal identity consists in degrees of psychological and/or physical continuity. So the rationale for thinking that it is possible to become a posthuman presupposes the revisionary, reductive view of personal identity. The intuitive view and the revisionary view are incompatible. And so, if Parfit is right, the standard rationale for transhumanism, taken as a whole, is incoherent.

The problem described here tends to be brushed aside in the literature on transhumanism. I think that this is a serious oversight. For it seems to me that Parfit's arguments are fatal to transhumanism as it is usually presented. To substantiate this claim, it is necessary to consider how transhumanists might respond.

⁸ In Weir ([forthcoming](#)) I argue that living indefinitely as a posthuman is not desirable. But my argument is based on assumptions that many will question.

Transhumanists who are determined to maintain the standard rationale for transhumanism face a dilemma. They must either: (i) stick with the revisionary view of personal identity and explain why one should value one's future existence; or (ii) revert to the intuitive view and explain why one should expect the person who wakes up after neural prosthesis or uploading to be oneself. If neither of these responses can be made to work, there is also a fall-back response (iii): abandon the standard rationale for transhumanism for one that does not say that transhumanism is a good idea because *we personally* will get to escape death and enjoy the posthuman world. In the following three sections, I consider each of these responses in turn. I argue that the obvious strategies for defending each response are either unpromising or mean changing transhumanism in ways that make it much less radical.

5. Response (i)

The first way in which transhumanists might respond to the problem presented in section four is by sticking with the revisionary view of personal identity and explaining why one should nonetheless value one's future existence. This will mean rejecting Parfit's contention that if the revisionary view is correct then personal identity does not matter. There are two ways in which transhumanists who adopt this response might proceed. They can:

- (i.i) attempt to refute Parfit's arguments that if personal identity is non-branching psychological continuity then one should not value one's future existence; or
- (i.ii) argue for some alternative version of the revisionary view that gets around Parfit's arguments.

Both options have been tried out in the literature on personal identity.

An influential argument that might be advanced in support of option (i.i) says that Parfit commits a fallacy in reasoning from the unimportance of the analysans, non-branching psychological continuity, to that of the analysandum, personal identity. Instead, so the argument goes, Parfit should have reasoned in the opposite direction, from the importance of personal identity to that of non-branching psychological continuity (Sosa, 2003, pp. 199–215; cf. Johnston, 2003, pp. 260–91).

Ernest Sosa backs up this argument with an analogy. Imagine someone who values cubes but who is indifferent to properties such as square-facedness and six-sidedness. One day, Sosa's cube

The Personal Identity Dilemma for Transhumanism

enthusiast discovers that cubes just are square-faced, six-sided solids. On doing so, Sosa (2003, pp. 214–15) argues, this person should not stop valuing cubes, they should start valuing square-faced, six-sided solids. Likewise, Sosa urges, on discovering that personal identity is non-branching psychological continuity, we should not stop valuing personal identity, we should start valuing non-branching psychological continuity.

Sosa argues that one should value non-branching psychological continuity even in cases like the double-upload scenario where, according to Parfit, doing so is absurd. Transhumanists who want to defend option (i.i) while avoiding this consequence might appeal to a similar response to Parfit put forward by Mark Johnston (2003).

Like Sosa, Johnson argues that Parfit should have reasoned from the importance of personal identity to that of non-branching psychological continuity, rather than in the opposite direction. But Johnston does not insist that one should value non-branching psychological continuity in cases of branching like the double-upload scenario. According to Johnston, in such cases it is reasonable to ‘extend your self-concern’ to future persons other than you, such as the uploads on Computer 1 and Computer 2. However, this ‘is not because identity is never what matters’, it is ‘because caring in this way represents a reasonable extension of self-concern in a bizarre case’ (Johnston, 2003, p. 282).

Neither defence of option (i.i) is promising. The problem with the claim that Parfit commits a fallacy in reasoning from the unimportance of the analysans, non-branching psychological continuity, to that of the analysandum, personal identity, is that the psychological continuity theory is not an analysis of personal identity as we intuitively understand it, but a *revisionary view* about its nature. So, the question is not whether one can reason from the unimportance of an analysans to that of the analysandum but whether, on revising our understanding of the nature of something, we can reasonably change our minds about its value. The answer is that we can.

To see this, suppose that you value amethyst for its magical power of preventing inebriation, but that you are indifferent to properties like being violet, and having the physical composition of quartz. One day you discover that amethyst is just a violet variety of ordinary non-magical quartz. Clearly, it is reasonable under these circumstances, to reassess the value of amethyst in light of your new understanding of its nature. And if previously it would have seemed absurd to you to value ordinary quartz for magical powers that it does not have, you can now extend this judgment to amethyst.

It is hard to imagine how one could go about arguing for the importance of non-branching psychological continuity, except by saying that we should transfer to it the value that we ordinarily place on personal identity. For this reason, the problem raised here is likely to apply to any version of option (i.i).

Transhumanists who adopt the kind of position defended by Sosa will face the additional problem of explaining away the powerful intuition that it is absurd to value non-branching psychological continuity in cases of branching. If having your psychological states uploaded onto Computer 1 is as good as ordinary survival, and having your psychological states uploaded onto Computer 2 is as good as ordinary survival, it seems very odd that having your psychological states uploaded onto Computer 1 and Computer 2 should be as bad as death. As Parfit (1984, p. 256) asks, 'how can a double success be a failure?' Instead, it seems that a transhumanist who adopts the revisionary theory should say that branching is no catastrophe, and certainly not a matter of life or death as we usually think of such matters.

Transhumanists who favour the kind of view defended by Johnston will avoid this difficulty. For Johnston accepts that on the revisionary view, branching is not a catastrophe. But transhumanists who adopt Johnston's position will face a different problem. For Johnston (2003, p. 282) gets around the absurdity of treating branching as a catastrophe only because he accepts that Parfit's arguments do indeed show that personal identity does not matter 'in certain bizarre cases which may never in fact arise'. Johnston merely urges that we do not generalise Parfit's conclusion beyond those cases. Transhumanists, however, are concerned with exactly the kind of bizarre cases to which Johnston is referring.

Suppose, for example, that you are considering whether to have your psychological states uploaded once or multiple times. A transhumanist who adopts the kind of view defended by Johnston ought to advise that, although having one's psychological states uploaded multiple times will end one's existence, one should not regard this fact as a significant drawback. It is hard to see how a transhumanist who is ready to give such advice could maintain the rationale for transhumanism described above, which depends on the assumption that one has a significant interest in one's future existence.

Suppose that option (i.i) is indeed unsatisfactory. Transhumanists might hope that the problem lies in Parfit's account of personal identity as non-branching psychological continuity. If so, there might be some superior version of the revisionary view, as per option (i.ii), that

The Personal Identity Dilemma for Transhumanism

also allows persons to survive enhancements such as neural prosthesis and uploading, and on which it makes sense to value one's future existence.

Transhumanists who defend option (i.ii) need a theory of personal identity that is sufficiently close to the psychological continuity theory to make it plausible that persons will survive radical enhancements such as neural prosthesis and uploading, but that gets around Parfit's arguments for thinking that personal identity does not matter. An obvious candidate is David Lewis's (1983) four-dimensional variant on the psychological continuity theory.

Lewis agrees with Parfit that personal identity consists in psychological continuity. But Lewis rejects Parfit's 'non-branching' clause. Instead, he represents persons as aggregates of temporally extended 'person-stages'. This allows Lewis to say that in cases of branching, there have existed *two persons* all along, and both survive the procedure. For example, in the double-upload scenario, there are two persons who start out sharing the same person-stage but end up with distinct person stages located on different computers.

Lewis claims that his theory vindicates the common-sense assumption that one has a significant interest in one's future existence. For even Parfit grants that one can place some value on the psychological continuities in which personal identity consists. And once the non-branching clause has been dropped, Lewis suggests, this will come to the *same thing* as valuing one's future existence. For example, Lewis can say that one should be happy to undergo double-uploading, not because personal identity does not matter, but because in the double-upload scenario, both persons who share one's current person-stage survive.

Lewis's variant on the psychological continuity theory has the surprising consequence that there might be countless distinct persons collocated with oneself right now, ready to divide on some future occasion. For this reason, a defence of option (i.ii) that appeals to Lewis's theory will reduce the intuitive appeal of transhumanism. But transhumanists who sympathise with Lewis's theory might see this cost as acceptable, or at any rate inevitable.

The main problem with Lewis's theory, in the present context, has been pointed out by Parfit (1976, pp. 74–75) himself. Despite what Lewis says, it is not true that on his theory valuing psychological continuity comes to the same thing as valuing one's future existence.

To see this, consider the variant on the double-upload scenario where Computer 2 is destroyed at the end. According to Lewis, this scenario involves two persons who share a person-stage prior to uploading. Initially both survive, but Person 1 goes on living

whereas Person 2 ceases to exist when Computer 2 is destroyed. Clearly Person 2 ought to feel differently about this new scenario, depending on whether they care about psychological continuity or personal identity. For after Computer 2 has been destroyed, there will exist someone who is psychologically continuous with Person 2, but no Person 2.

In his response to Parfit, Lewis (1983, pp. 73–77) concedes that according to his theory, in scenarios like this, one can only rationally desire that *some* future person should be psychologically continuous with oneself, not that that person should *be oneself*. So Lewis's theory does not vindicate the common-sense assumption that one has a significant interest in one's future existence after all.

In fact, even if we set aside cases of branching, it is clear that Lewis's theory fails to justify the thesis that one has a significant interest in one's future existence. For if personal identity consists in degrees of psychological continuity then the facts about personal identity depend on how much continuity we demand before we count someone as 'the same person'. As Lewis (1983, p. 70) acknowledges, 'the choice of this cutoff point is more or less arbitrary'. If so, then it seems that whether one survives relative to a given specification of the requisite degree of continuity can have no practical importance.

For both reasons, an attempt to defend option (i.ii) by appeal to Lewis's theory is unlikely to succeed. Transhumanists who adopt option (i.ii) must identify some other version of the revisionary view of personal identity that allows persons to survive radical enhancements such as neural prosthesis and uploading, and on which one has a significant interest in one's future existence. I am not aware of any more promising candidate.⁹

A final, general point about response (i). Suppose some version of the revisionary view can be formulated on which personal identity

⁹ Two ideas that deserve mention here are Michael Cerullo's (2015) theory of branching identity, and the thesis that persons are repeatable types rather than particular tokens, which receives a somewhat sympathetic treatment by Walker (2014). These have been subjected to persuasive criticism by Bauer (2017) and by Schneider (2019, pp. 130–34) and Goldwater (2021) respectively, however. Both proposals allow two distinct future persons to be oneself. The basic problem with such theories is that particulars are meant to obey the indiscernibility of identicals. Since two distinct future particulars will *ex hypothesi* be discernible, they cannot be numerically identical. As a result, theories that allow two future persons to be oneself must either reject the indiscernibility of identicals or treat persons as universals rather than particulars. Both options are counterintuitive.

The Personal Identity Dilemma for Transhumanism

consists in X continuities, qualified in Y way, and that it makes sense to place some value on the existence of these continuities, so qualified. Even so, this will not be *the same sort of value* that it would make sense to place on one's future existence if the intuitive view of personal identity were correct. For as Parfit observes, the value that it makes sense to place on the continuities is likely to be comparatively small, and to diminish as the continuities grow weaker.

So, although such a theory might allow us to *say* that the merit of transhumanism consists in the opportunity to avoid death and go on living, we will not *mean* by this what a hearer uninitiated in revisionary theories of personal identity would take us to mean. And we should be obliged, for the sake of clarity, to stress that *all* we mean by this is that radical enhancements will preserve such and such continuities in such and such a way.

Two consequences follow. First, the rationale for transhumanism, so clarified, is likely to have a comparatively narrow appeal. For, many of us will have little interest in the existence of future persons who have memories of our experiences, intentions to carry out our plans etc. if those persons are not in any deeper sense *us*. I, at any rate, would rather live on through ordinary means like having children and writing philosophy articles. Secondly, those who do place a significant degree of value on the continuities will have correspondingly strong grounds to favour enhancements that maximise the strength of the continuities. This will mean favouring enhancements that are *conservative* about human nature.

Even in the best-case scenario, therefore, it seems that option (i) will result in a rationale for a comparatively small class of persons to undergo enhancements that avoid bringing about major changes in human nature. It might be possible to defend a version of transhumanism on this basis, but the resulting position will lack the broad appeal and radical nature that transhumanism is ordinarily supposed to have.

6. Response (ii)

The problem with response (i) is that it is hard to explain why one should place a significant degree of value on one's future existence if the revisionary view of personal identity is correct. Parfit's arguments to the contrary are powerful and merit their massive influence in the literature. For this reason, it might be easier to defend transhumanism by adopting response (ii), that of reverting to the intuitive, nonreductive view of personal identity, and explaining why one

should expect the person who wakes up after neural prosthesis or uploading to be oneself.

The intuitive view of personal identity has two important advantages for transhumanists. First, because it says that personal identity is a further fact over and above one's physical and psychological properties, the intuitive view places no limit on the transformations that one could, in principle, survive. Secondly, it is widely accepted that if the intuitive view is correct, then one has the kind of interest in one's future existence that makes death a great evil and survival a genuine good. So transhumanists who adopt option (ii) need not worry about arguing that one should value one's future existence.

Transhumanists who adopt option (ii) must, however, provide an argument for thinking that the person who wakes up after radical enhancements such as neural prosthesis or uploading will be oneself, and one that does not rest on the assumption that personal identity consists in degrees of psychological and/or physical continuity. Although some proponents of the intuitive view of personal identity have expressed sympathy with transhumanism (e.g., Göcke 2017), there has not been much discussion of whether such an argument can be given.

One exception is an argument advanced by David Chalmers (2014, pp. 111–13) which is compatible with the intuitive view of personal identity, though it does not entail it, and which transhumanists might therefore employ in defence of response (ii). Chalmers asks us to imagine a process of 'gradual uploading'. First, 1% of your brain is replaced by a silicon circuit that performs the same function. Then another 1%, and so on until your entire brain has been replaced. (Again, it is not obvious that it is possible to reproduce the functions of the nervous system with silicon circuits, or that if this is possible, the result will be a conscious person. But, I grant these assumptions here, in order to focus on the issue of personal identity.) Chalmers (2014, p. 111) suggests that you would survive the first step of this process. For if not, 'this would raise the possibility that everyday neural death may be killing us without our knowing it'. Chalmers reasons that if you would survive the first step, then you would survive each subsequent step too, for they are effectively identical. If so, then the person with a 100% artificial brain must be you.

Chalmers (2014, p. 112) reinforces this conclusion in two ways. First, 'if gradual uploading happens, most people will become convinced that it is a form of survival'. For 'it will be very unnatural for most people to believe that their friends and families are being killed by the process'. Secondly, if silicon circuits that perform the

The Personal Identity Dilemma for Transhumanism

same functions as the nervous tissue they replace support consciousness, then there will be a 'continuous' stream of consciousness throughout the process. Chalmers suggests that if so, the same person must survive throughout.

Having argued that you would survive 'gradual uploading', Chalmers (2014, p. 113) adds that we can imagine speeding up the process, so that it only takes minutes or seconds. If you survive when the process is carried out slowly, it seems that you will also do so when it is carried out quickly, or even instantaneously. Hence, Chalmers concludes that if you would survive 'gradual uploading' then you would survive instantaneous uploading too.

'Gradual uploading' is essentially what others have called 'neuron replacement therapy' or 'neural prosthesis'. So if Chalmers' argument is successful, then one can be confident that the person who wakes up after neural prosthesis or uploading will be oneself, even on the intuitive, nonreductive view of personal identity. Transhumanists who adopt the intuitive view of personal identity will then be in a position to claim that, if one undergoes the necessary enhancements, some future posthuman will *really be oneself*, in a sense worth caring about deeply. This would vindicate the standard rationale for transhumanism.

Taken by itself, Chalmers' argument provides a *prima facie* plausible case for thinking that on the intuitive view of personal identity, one would survive neural prosthesis or uploading. Four objections can be raised, however, which suggest that a defence of response (ii) that appeals to Chalmers' argument will be of little or no help to transhumanists.

First, as Chalmers (2014, pp. 109–11) himself observes, his argument is counterbalanced by a comparably strong counterargument. The counterargument starts with the premiss that if a digital replica of one's mind were created, while the original still exists, that replica would not be oneself. If so, it continues, then if a digital replica of one's mind were created and the original *destroyed*, the replica still would not be oneself. But enhancements such as neural prosthesis and uploading just consist in the creation of a digital replica of one's mind and the destruction of the original. And so it seems that one would not survive neural prosthesis or uploading after all.

This counterargument assumes that the destruction of the original will not affect whether a digital replica of one's mind is oneself. This assumption would be false on many versions of the revisionary view of personal identity. But it would be accepted by most proponents of the intuitive view. For, on the intuitive view, personal identity seems to involve something like an indivisible soul. And although it *could* be

the case that when the original is destroyed, one's soul hops over to the replica, we have no reason to believe that this is so.

Secondly, Chalmers' rationale for thinking that one could survive every step in the process of gradual uploading contains an invalid inference. Chalmers (2014, p. 112) suggests that one would survive having the first 1% of one's brain replaced with a silicon circuit on the basis that otherwise 'everyday neural death may be killing us without our knowing it'. This seems reasonable. But it does not follow that one would survive having the ninetieth or hundredth percent of one's brain replaced by silicon circuits. After all, I think I could survive having 1% of my brain replaced by a Lego brick. For otherwise fairly commonplace head injuries might be killing us without our knowing it. But it would be wrong to infer that I could survive with a brain composed ninety or one hundred percent of Lego.

Of course, Chalmers' silicon circuits are meant to reproduce the functional properties of the nervous tissue they replace, while Lego bricks do not. But this makes no difference unless we assume that functional continuity preserves personal identity. Such an assumption would be question-begging. For it is exactly what Chalmers' argument is meant to prove.

Joseph Corabi and Susan Schneider (2012) raise two further problems for Chalmers' argument that you would survive 'gradual uploading'. First, Chalmers appears to assume that if your friends and family would regard you as the same person after 'gradual uploading', this would make it more probable that the upload is you. This assumption is implausible. For your friends and family would respond in the same way if you were replaced by a replica who they are unable to distinguish from you.

Second, Chalmers' appeal to the continuous stream of consciousness that exists throughout 'gradual uploading' seems to be question-begging. For if 'continuous' means something sufficient for the preservation of personal identity then it does not follow from the thesis that the silicon circuits used in 'gradual uploading' support consciousness that there is a *continuous* stream of consciousness throughout. It could be the case that, although there exists conscious experience at every moment in the process, in the early stages this is the conscious experience of the original person, whereas at later stages it is the experience of a new, artificial intelligence, that has replaced the original person. But if 'continuous' means something insufficient for the preservation of personal identity then it does not follow from the thesis that there is a continuous stream of consciousness that personal identity is preserved.

The Personal Identity Dilemma for Transhumanism

For these reasons, Chalmers' argument does not provide compelling grounds for thinking that, on the intuitive view of personal identity, one would survive neural prosthesis or uploading. Taken as a whole, Chalmers' discussion merely illustrates the fact that if the intuitive view is correct, then it is difficult to judge which transformations persons can and cannot survive.

Some transhumanists propose that superintelligent AI will solve this problem by telling us which processes do and which do not preserve personal identity (e.g., Bostrom, 2014, pp. 245–6; Turchin and Chernyakov 2024). But it is not clear how superintelligent AI is supposed to help. For there is no reason why any amount of computational power should be able to determine facts about personal identity on the basis of facts about physical and psychological properties of which personal identity is, on the intuitive view, logically independent. At face value, what is needed is not more computational power but new *data* concerning the toing and froing of souls (or irreducible personal-identity facts, or whatever it is that personal identity, on the best version of the intuitive view, involves).

It is worth noting the contrast between the position of transhumanists and that of traditional theists in this respect. Both can appeal to the intuitive view of personal identity to explain why we should want to get to the world to come. But theists can rely on God to get us there, whereas transhumanists must rely on technologies based on natural science. This is a problem for transhumanists. For the intuitive view makes the facts about personal identity highly mysterious, and not obviously amenable to the methods of natural science.

In the absence of some means of tracking souls, the only obvious way for proponents of the intuitive view of personal identity to be confident that persons will survive a given enhancement is by making that enhancement as moderate as possible, so as to avoid accidentally replacing the original person with a replica. For this reason, transhumanists who adopt response (ii), like those who adopt response (i), ought to favour enhancements that are as conservative as possible about human nature. Again, it might be possible to defend a version of transhumanism on this basis, but it will be much less radical than transhumanism as it is ordinarily presented.

7. Response (iii)

It is not easy to defend the standard rationale for transhumanism on either the intuitive or the revisionary view of personal identity. As I mentioned in section four, there is also a fall-back response (iii), that

of abandoning the standard rationale for transhumanism in favour of one that does not depend on the value of one's personal survival in the posthuman world. I have called this a 'fall-back' response because it concedes that the standard rationale for transhumanism is not successful. But in view of the problems raised for responses (i) and (ii), response (iii) might be the best way out of the personal-identity dilemma.

Transhumanists who adopt response (iii) must provide a rationale for transhumanism that does not depend on the value of one's personal survival in the posthuman world. An obvious strategy for providing such a rationale is to appeal to Parfit's contention that although personal identity does not itself matter, the continuities in which it consists do. For if so, it ought to be possible to motivate enhancements such as neural prosthesis and uploading on the basis that they preserve the relevant continuities, irrespective of whether they preserve personal identity. This is the kind of view that Bostrom appears to have in mind when he says:

Preservation of personal identity, especially if this notion is given a narrow construal, is not everything. We can value other things than ourselves, or we might regard it as satisfactory if some parts or aspects of ourselves survive and flourish, even if that entails giving up some parts of ourselves such that we no longer count as being the same person. (Bostrom, 2005, p. 9)

Though Bostrom does not develop the point in detail, this passage suggests that it might be possible to motivate transhumanism on purely Parfitian grounds.

Bostrom makes this point in passing, as if it need not entail any major departure from the standard rationale for transhumanism. But this is misleading. A defence of response (iii) that appeals to the kind of position defended by Parfit will contrast dramatically with most transhumanist literature, including texts authored by Bostrom himself. To see this, consider once again Bostrom's 'Letter from Utopia.' The 'Letter' is signed off 'Your Possible Future Self' and the writer emphasises that if the transhumanist project succeeds

then I am not just a possible future, but your actual future ... I am writing to tell you about my life—how marvellous it is—that you may choose it for yourself. (Bostrom, 2008, pp. 1, 7)

Passages like this are designed to appeal to the kind of self-interested concern that, according to Parfit, only makes sense on the intuitive view of personal identity. Transhumanists who adopt the kind of position defended by Parfit must give up on this kind of rhetoric.

The Personal Identity Dilemma for Transhumanism

The same argument applies to passages that appeal to the tragedy of death. As we saw above, this is a frequent point of emphasis in transhumanist literature. Elizar Yudkowski (2024, Ch. 45, pt. 3) goes so far as to suggest that ‘the descendants of humanity’ will ‘weep to hear that such a thing as Death had ever once existed.’ But transhumanists who adopt the kind of view defended by Parfit should view the idea that death is a great tragedy as mistaken. For if Parfit is right then although death

can seem depressing [...] the reality is only this. After a certain time, none of the thoughts and experiences that occur will be directly causally related to this brain, or be connected in certain ways to these present experiences. That is all this fact involves. And, in that redescription, my death seems to disappear. (Parfit, 2001, p. 33)

This could hardly be further from the superlatively negative view of death that is typically presupposed by authors pressing the case in favour of transhumanism.

A defence of response (iii) that appeals to the kind of position defended by Parfit must, then, abandon a great deal of what is ordinarily said in favour of transhumanism. It should focus instead on the idea that psychological and/or physical continuity is to be valued in itself, and that this alone is enough to make radical enhancements such as neural prosthesis and uploading worthwhile.

It is plausible that a coherent rationale in favour of at least some kinds of enhancement can be made on this basis. And transhumanists who rest their case solely on such a rationale will not have to worry about the dilemma presented in this paper. To that extent, a version of response (iii) that appeals to the kind of position defended by Parfit might seem attractive. There are two problems with this kind of response, however, in view of which it is likely to be no more satisfactory than responses (i) or (ii). Both problems were already raised in the context of response (i) above.

First, many of us will have little interest in the existence of future persons who are psychologically and/or physically continuous with us, but who are not in any deeper sense *us*. For my own part, I see no appeal in the idea of bequeathing one or more digital replicas of myself to posterity if none of them is actually me. If anything, I find the idea unsettling. The widespread focus on the desirability of survival and the evil of death in the literature on transhumanism suggests that many transhumanists will feel the same way.

Second, those who do value mere psychological and/or physical continuity will have correspondingly strong grounds to favour

enhancements that maximise the continuities. This will mean favouring enhancements that change as little as possible about the relevant psychological and/or physical characteristics. So, as with responses (i) and (ii), a defence of response (iii) based on the kind of view defended by Parfit is also likely to result in a position much less radical than transhumanism as it is usually understood.

It is worth adding that it is not even clear that the technologically enhanced persons who will someday exist on this version of transhumanism will qualify as ‘transhumans’ or ‘posthumans’. For as these terms are usually used, a ‘transhuman’ or ‘posthuman’ is supposed to be someone who was *formerly a human*. But proponents of response (iii) are not interested in ensuring that human persons survive as transhumans or posthumans, only in producing enhanced persons who stand in certain continuity relations with human persons.

In any case, a defence of response (iii) that appeals to the kind of view defended by Parfit is likely to result in a position that lacks the broad appeal and radical nature that transhumanism is usually supposed to have. Perhaps a more promising strategy for defending response (iii) can be found. But I leave it to proponents of transhumanism to pursue the question further.¹⁰

8. Conclusion

In explaining why one should want to become a posthuman, transhumanists appeal to the kind of self-concern that, according to Parfit, only makes sense if the intuitive, nonreductive view of personal identity is correct. But in explaining how it is possible to become a posthuman they appeal to the revisionary, reductive view of personal identity. So if Parfit is right, transhumanists have committed a kind of philosophical bait-and-switch. The rationale for becoming a posthuman and the rationale for thinking that it is possible to do so presuppose incompatible views of personal identity. The obvious responses are either unsatisfactory or mean rethinking transhumanism in ways that make it much less radical. Parfit’s arguments appear to be fatal to transhumanism as it is widely understood.¹¹

¹⁰ One theorist who has argued that transhumanists should dispense with the idea that radical enhancements will preserve personal identity is James Hughes (2013). While Hughes’ position seems vulnerable to the objections raised here, it deserves to be discussed in detail elsewhere.

¹¹ I am very grateful to Andrew Pinsent, Mikolaj Slawkowski-Rode, and two anonymous reviewers for *Philosophy* for comments on a draft of

The Personal Identity Dilemma for Transhumanism

References

- William Bauer, 'Against Branching Identity', *Philosophia*, 45 (2017), 1709–19.
- Elise Bohan, *Future Superhuman* (Sydney: University of New South Wales Press, 2022).
- Nick Bostrom, 'Transhumanist Values', in Frederick Adams (ed.), *Ethical Issues for the 21st Century* (Charlottesville, VA: Philosophical Documentation Center Press, 2005), 3–14.
- Nick Bostrom, 'Letter from Utopia', *Studies in Ethics, Law, and Technology*, 2:1 (2008), 1–7.
- Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014).
- Joseph Butler, *The Analogy of Religion*, David McNaughton (ed.), (Oxford: Oxford University Press, 2021 [1736]).
- Michael Cerullo, 'Uploading and Branching Identity', *Minds & Machines*, 25 (2015), 17–36.
- David Chalmers, 'Uploading: A Philosophical Analysis', in Russell Blackford and Damien Broderick (eds.), *Intelligence Unbound: The Future of Uploaded and Machine Minds* (Oxford: Wiley Blackwell, 2014), 102–118.
- Joseph Corabi and Susan Schneider, 'The Metaphysics of Uploading', *Journal of Consciousness Studies*, 19:7–8 (2012), 26–44.
- Aubrey D. N. J. de Grey, 'Life Span Extension Research and Public Debate: Societal Considerations', *Studies in Ethics, Law, and Technology*, 1:1 (2007), 1–13.
- Benedikt Paul Göcke, 'Christian Cyborgs: A Plea for a Moderate Transhumanism', *Faith and Philosophy*, 34:3 (2017) 347–64.
- Jonah Goldwater, 'Uploads, Faxes, and You: Can Personal Identity Be Transmitted?', *American Philosophical Quarterly*, 58:3 (2021), 233–50.
- William Hazlitt, *An Essay on the Principles of Human Action and Some Remarks on the Systems of Hartley and Helvetius*, reprinted with an introduction by J. R. Nabholz (Gainesville, FL: Scholar's Facimiles & Reprints, 1969 [1805]).
- James Hughes, 'Transhumanism and Personal Identity', in Max More and Natasha Vita-More (eds.), *The Transhumanist Reader* (Oxford: Wiley-Blackwell, 2013), 227–34.

this paper, and to Alex McLaughlin, Parker Settecase, and a seminar group at the University of Warsaw for helpful discussions. This research was supported by the University of Oxford project 'New Horizons for Science and Religion in Central and Eastern Europe' funded by the John Templeton Foundation. The opinions expressed in the publication are those of the author and do not necessarily reflect the view of the John Templeton Foundation.

Ralph Stefan Weir

- Humanity+, *The Transhumanist FAQ: v3.0*, accessed 25 January 2024, <https://www.humanityplus.org/transhumanist-faq>.
- Mark Johnston, 'Human Concerns without Superlative Selves', in Raymond Martin and John Barresi (eds.), *Personal Identity* (Oxford: Blackwell, 2003), 260–91.
- Ray Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (New York: Viking, 1999).
- Ray Kurzweil, *The Singularity is Near* (New York: Viking, 2005).
- Ray Kurzweil and Kathleen Miles, 'Nanobots in Our Brains Will Make us Godlike', *New Perspectives Quarterly*, 32:4 (2015), 24–29.
- David Lewis, 'Survival and Identity', in David Lewis, *Philosophical Papers Vol. 1* (Oxford: Oxford University Press, 1983), 55–77.
- Watanabe Masataka, *From Biological to Artificial Consciousness: Neuroscientific Insights and Progress* (Cham: Springer, 2022).
- Eric Olsen, 'The Central Dogma of Transhumanism', in Boran Berčić (ed.), *Perspectives on the Self* (Rijeka: University of Rijeka Press, 2017), 35–57.
- Derek Parfit, 'Personal Identity', *The Philosophical Review*, 80:1 (1971), 3–27.
- Derek Parfit, 'Lewis and Perry and What Matters', in Amélie Rorty (ed.), *The Identities of Persons* (London: University of California Press, 1976), 91–107.
- Derek Parfit, 'Personal Identity and Rationality', *Synthese*, 53:2 (1982), 227–41.
- Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984).
- Derek Parfit, 'The Unimportance of Identity', in Henry Harris (ed.), *Identity* (Oxford: Oxford University Press, 2001), 13–45.
- Massimo Pigliucci, 'Mind Uploading: A Philosophical Counter-Analysis', in Russell Blackford and Damien Broderick (eds.), *Intelligence Unbound: The Future of Uploaded and Machine Minds* (Oxford: Wiley Blackwell, 2014), 119–30.
- Adam Piore, 'The Neuroscientist Who Wants To Upload Humanity To A Computer', *Popular Science*, 16 May 2014.
- Susan Schneider, *Artificial You: AI and the Future of Your Mind* (Oxford: Princeton University Press, 2019).
- Ernest Sosa, 'Surviving Matters', in Raymond Martin and John Barresi (eds.), *Personal Identity* (Oxford: Blackwell, 2003), 199–215.
- Liz Stillwaggon Swan and Joshua Howard, 'Digital Immortality: Self or 0010110?', *International Journal of Machine Consciousness*, 4:1 (2021), 245–56.
- Richard Swinburne, 'Personal Identity', *Proceedings of the Aristotelian Society*, 74 (1973–1974), 231–47.
- Alexey Turchin and Maxim Chernyakov, *Classification of Approaches to Technological Revival of the Dead*, accessed 25 January 2024, <https://philpapers.org/archive/TURCOA-3%20.pdf>
- Mark Walker, 'Uploading and Personal Identity', in Russell Blackford and Damien Broderick (eds.), *Intelligence Unbound: The Future of Uploaded and Machine Minds* (Oxford: Wiley Blackwell, 2014), 161–77.

The Personal Identity Dilemma for Transhumanism

Ralph Stefan Weir, 'Transhumanismus und die Metaphysik der menschlichen Person', in Benedikt Paul Göcke and Frank Meier-Hamidi (eds.), *Designobjekt Mensch: Die Agenda des Transhumanismus auf dem Prüfstand* (Freiburg im Breisgau: Herder, 2018), 225–58.

Ralph Stefan Weir, 'The Logical Inconsistency of Transhumanism', *Philosophy, Theology and the Sciences*, 10:2 (2024), 199–220.

Ralph Stefan Weir, 'Transhumanist Immortality is Neither Probable nor Desirable', in Daniel Came and Stephen Burwood (eds.), *Transhumanism and Immortality* (Budapest: Trivent, forthcoming).

Eliezer Yudkowsky, *Harry Potter and the Methods of Rationality*, accessed 25 January 2024, https://www.fanfiction.net/s/5782108/1/Harry_Potter_and_the_Methods_of_Rationality.

RALPH STEFAN WEIR (rweir@lincoln.ac.uk) is Senior Lecturer in Philosophy at the University of Lincoln. His recent publications include *The Mind-Body Problem and Metaphysics* (Routledge, 2023) and *From Existentialism to Metaphysics: The Philosophy of Stephen Priest* (Peter Lang, 2021).