THE FALLACY OF FAVORING GRADUAL REPLACEMENT MIND UPLOADING OVER SCAN-AND-COPY

Keith B. Wiley, PhD¹ and Randal A. Koene, PhD²

¹Fellow, Brain Preservation Foundation
Author of A Taxonomy and Metaphysics of Mind-Uploading
kwiley@keithwiley.com
http://keithwiley.com, http://brainpreservation.org

²Founder, Carboncopies.org
randal.a.koene@carboncopies.org
http://carboncopies.org, http://minduploading.org

Abstract

Mind uploading speculation and debate often concludes that a procedure described as gradual in-place replacement preserves personal identity while a procedure described as destructive scan-and-copy produces some other identity in the target substrate such that personal identity is lost along with the biological brain. This paper demonstrates a chain of reasoning that establishes metaphysical equivalence between these two methods in terms of preserving personal identity.

Keywords: mind uploading, whole brain emulation, personal identity, metaphysics, scan, copy, replacement

Introduction

Newcomers to mind uploading philosophy [Strout 1997] quickly become familiar with a shared set of canonical thought experiments. Two of the most popular are gradual in-place replacement and destructive scan-and-copy. The former consists of steadily replacing individual components of the brain, say neurons, with microscopic devices of functional equivalence, while the latter stabilizes the brain via vitrification or plastination, then sections and scans the static brain, and then instantiates the scan via whole brain emulation (WBE) in a computational substrate [Sandberg & Bostrom 2008, Koene & Deca 2014].

A popular conclusion after considering these scenarios is to view gradual in-place replacement more favorably than scan-and-copy with regard to successfully preserving or transferring personal identity (sometimes called consciousness, a shorthand for phenomenal consciousness, or our ongoing experience of ourselves) [Block 1995, Chalmers 2010, Olson 2010]. Gradual in-place replacement is granted the status of successful identity preservation while scan-and-copy is denigrated as producing a copy, the implication being that this is some other person, the mind entering the procedure having died [Corabi & Schneider 2012, Dorrier 2015, Hay 2014, Josh 2009, Morris 2013, Schneider 2014]. This article proposes that contrary to these opposing interpretations, both scenarios should actually be considered metaphysically similar in the ultimate identity status of the minds they produce, either successful transfer preserving personal identity or failure producing some other. The argument is two-part, first equating slow gradual in-place replacement (SGR) with quick gradual replacement (QGR) or even instantaneous replacement (IR) via an argument of temporal comparisons, and then equating IR with destructive scan-and-copy (SaC) via an argument of spatial comparisons. This reasoning establishes a transitive relation connecting SGR to SaC using IR as an intermediate stage. The implication of this transitive equivalence is that all considered procedures should receive the same identity status, ostensibly successful transfer preserving personal identity.

Throughout this paper we refer to two alternative interpretations: the preservation of personal identity by the new substrate versus the emergence of some other identity. The terms transfer and copy sometimes indicate these two cases. Transfer implies associating one’s personal identity with the new substrate. Copy implies metaphysical failure of that association, so that some other identity is created and associated with the
new substrate instead; *copy* further implies that since the biological brain is destroyed by both SGR and SaC, personal identity is presumed lost. The term *transfer* can be confusing when discussing SGR due to its overly spatial implications, so we prefer to speak of *preserving* personal identity in the new substrate. Similarly, the term *copy* is confusing when implying failed identity preservation because it is broadly used in other ways, e.g., copying a mind, personal identity, neural functionality, cognitive state (knowledge, memories, personality, etc.), brain scan data, physical brain structure or material pattern, etc. It appears in the popular label for the mind uploading method known as *scan-and-copy*, but there expresses purely technical aspects of data copying and does not speak to identity preservation one way or the other. For clarity, we avoid the term *copy* as used in the claim *it's-just-a-copy*, and instead consider whether personal identity is *preserved* when uploaded into a new substrate or whether some *other* identity emerges.

**Equating Slow Replacement with Quick Replacement**

Variants of the temporal argument have been proposed before, such as Chalmers’ chapter in Blackford & Broderick [2014]. This argument considers a spectrum of temporal rates of gradual replacement, measured in neurons replaced per second or some comparable metric. Scenario 1.1.1.1.1 of the taxonomy in Wiley [2014] offers a detailed description, which incidentally also shows that initial expectations that SGR is positively languid may not agree with replacement rates required for a procedure of reasonable expediency. For example, a total procedure time of 100 days, operating around the clock, would require replacing 10,000 neurons every second nonstop. To complete the procedure in a twenty-four hour period would require continually replacing one million neurons every second for an entire day. Putting aside whether a procedure of practical duration would actually be meaningfully *slow*, incremental replacement is nevertheless how gradual replacement is generally presented and in this way it differs from SaC, which equates to global instantaneous (or at least discontinuous) identity transfer. At faster replacement rates, SGR approaches QGR or even IR, thereby resembling SaC in some crucial aspects.

Some readers may be inclined to assume that the fast end of the spectrum would lead to a failed transfer, with some other person resulting from the upload process, while the one who entered the procedure died. For example, SGR is often hailed for maintaining some notion of a *stream of conscious continuity* [Chalmers 2010, Van Gulick 2014]). In *it's-just-a-copy* debates, arguments to the effect that nongradual procedures fail to transfer this stream are often raised. However, defending such a conclusion falls to its proponent—perhaps the alleged conscious stream can survive quick or instantaneous replacement after all, or perhaps the entire idea of a consciousness stream is simply flawed [Blackmore 2002, Dennett 1991]). For readers who are comfortable granting preserved identity to QGR and IR, but who nevertheless deny it to SaC, we address that view later, where we discuss the spatial relation between IR and SaC.

To hold the position that slow replacement preserves identity but that quick replacement does not implies that some success-to-failure transition must happen somewhere in the range of possible replacement rates. One might claim that there exists a cutoff along the replacement rate spectrum where personal identity preservation suddenly flips to producing some other identity. Immediately below the cutoff, the rate would be considered sufficiently slow to allow total preservation of personal identity, yet at an infinitesimally faster rate, the interpretation would be to utterly reject metaphysical transfer or preservation, resulting in some other identity instead. While all mental functionality would still technically transfer to the new substrate (this prerequisite is clarified below), thereby producing an identically functioning, thinking and behaving person (Chalmers’ *functional isomorph* [Chalmers 2010]), detractors would apparently claim that the crucial *properties of personal identity* have failed to ride along, and for no better reason than the minutest increase in replacement rate. Proposing such a cutoff requires not only a defense that it is a valid concept in the first place, but furthermore some notion of where on the replacement rate spectrum it actually resides.
Some readers might propose that the transition from identity preservation to other identity not necessarily conform to a discrete cutoff, but rather to a smooth continuum cross-dissolving from full preservation (SGR) to full other (QGR or IR) as the replacement rate increases, and with intermediate replacement rates resulting in a mixture of both. Theories of personal identity remain quite diverse (offering options such as bodily or biological identity, which ties identity to our bodies and possibly even requiring biological substrate, versus psychological identity, which ties identity to total memory state, plus other theories as well [Olson 2010]). Some theories would exclude identity preservation in any mind uploading scenario, such as Hopkins’ [2012] and Searle’s [1997] preference for biological identity. Others might support the proposal that certain uploading scenarios can yield blends of identity preservation and an invoked other identity. And yet other theories might be less forgiving, prescribing an all-or-nothing conclusion in any given scenario. However, there is a related concept, \textit{survival}, which is a judgement that places some outcomes into the survived category and some into the not survived category. Chalmers [2010] makes a similar distinction. On the question of whether the person entering an uploading procedure explicitly \textit{survives} to then associate with the new substrate, or dies with the biological substrate, we should draw a discrete conclusion in any given uploading scenario. Relevantly, in it’s-just-a-copy debates, and by the nature of the it’s-just-a-copy argument when dismissing the identity status of an upload, survival is actually the paramount concern, perhaps even more so than possibly nebulous identity. When we ask whether the person resulting from a destructive uploading procedure is the same person who entered the procedure, we are directly asking whether the latter survived. And when we denigrate the upload as being a mere copy, we are indicating in the same breath that the person who entered the procedure died—they did not survive. An intermediate conclusion on the question of survival quite simply betrays the meaning of the word. Consequently, we needn’t consider smooth transitions of identity in so far as such nuances do not extend to the underlying survival, \textit{the actual primary concern}. This leaves us faced with the same challenge previously raised: for there to be survival in SQR but death in QGR or IR requires a discrete cutoff of the most infinitesimal change in replacement rate somewhere in between, a seemingly unlikely proposition.

To clarify a possible point of confusion, we are not considering whether there are valid practical or strategic reasons to choose one procedure over another, nor are we considering whether gradual in-place replacement has a physiological and technical speed limit, i.e., whether replacement above a certain rate might yield an upload that fails to adequately preserve the original neural functionality and cognitive behavior and thereby fails to sufficiently resemble the person who entered the procedure. While the matter of a technical speed limit is doubtlessly crucial from an engineering perspective, it is irrelevant to this philosophical discussion \textit{precisely} because detractors often grant this concession when denying identity status to an uploaded individual. The spirit of the \textit{it’s-just-a-copy} claim is that \textit{even if perfect preservation of function is achieved}, we should nevertheless regard the person as dead and refuse to grant the uploaded mind the same personal identity, and by implication that we should similarly judge SaC. In this article, we are only considering whether \textit{metaphysical identity} is subject to some transition from preserved to other personal identity along the spectrum of replacement rates. Technical efficacy is off the table.

Since we are assuming technical efficacy, a transition of identity from preservation to other seems the unlikely state of affairs. Until a sufficient theory is presented as to why such a state-change should be expected, the default position should be that no such state-change occurs. Consequently, we should assign the same identity status at both ends of the spectrum. If we would grant identity status to a SGR individual, then we should grant that same status to a QGR individual, or even an IR individual, such as if all neurons are replaced simultaneously at the flip of a master switch.

If we conclude that no transition is ultimately defensible and we grant transfer status to an IR individual, then if there is also subsequently no meaningful difference between IR and SaC, we should grant the same status to SaC as well. This reasoning results in a conundrum for the position that SGR and
SaC should receive differing survival statuses. The realization of this inconsistency is the crux of this article. Having argued against the temporal distinction between SGR and SaC, the only remaining distinction is that SaC may involve a larger spatial translation from the biological brain to the new substrate (admittedly, procedures like the Moravec method blur this distinction [Moravec 1988, Wiley 2014]). In the next section we consider whether the spatial distinction between IR and SaC indicates any consequence for metaphysical identity or its survival, or whether they are fundamentally equivalent, in which case we should grant equivalent identity status across all three procedures.

Equating Quick Replacement with Destructive Scan-and-Copy

Some readers may feel that the distinction between IR and SaC lies in the likelihood that SaC involves a greater spatial translation of neural function from the biological brain to the new substrate, perhaps even on the belief that any in-place replacement procedure requires no spatial translation at all. In these debates it is sometimes asked how function and identity could possibly *move through space* from one brain to another. It may seem that minimizing or even negating any spatial translation of the neural functionality and metaphysical mind implies greater compatibility with identity preservation, especially as contrasted with spatially transferring a person’s mind and identity across a room to a cloned or robotic body. However, this position is difficult to rationalize.

First of all, the possible assumption that in-place replacement involves no spatial translation is simply incorrect. While scenario 1.1.1.1.2 in the taxonomy presented in Wiley [2014] offers a detailed explanation, the following summary is straightforward. There is always some spatial translation of function as a neuron *hands off* its neural behavioral role to a nearby microscopic device, and furthermore, the hand-off occurs discontinuously through space; this functionality does not move smoothly through space from location A to location B (in so far as we can speak of nonphysical functionality residing at a location in space or moving through it in the first place! Others would seem to agree [Hopkins 2012].). Rather, at one moment, some neural function is being performed by a biological neuron in one location; then at some later time that same function is being performed by a nearby artificial neuron in a spatially discontinuous and different location and the biological neuron is no longer operational. The two entities, a neuron and its replicant, cannot spatially coincide even if they are so near one another as to physically abut (~10–100 microns from center of neuron to center of replicant)—there is simply always some distance to be considered (to a committed functionalist, this consideration of the nuances of physical distance may sound tedious and irrational, for what possible relationship could there be between physical distance and nonphysical function, but we must consider such matters to complete the argument).

Recognizing that some spatial translation is always involved, we can consider whether there remains a fundamental difference in this translation between the two scenarios. In SaC, the distance over which this functionality purportedly transfers is on the order of meters at a minimum (from one body to another in the same room) or arbitrarily further. For example, teleportation, a variant of SaC, could transfer mental function any conceivable distance, including across the universe and/or far into the future. But as shown above, in-place replacement involves some spatial translation as well. Consequently, any spatial distinction must now hinge on distance comparisons. Curiously, Wiley [2014] shows that the micron-scale per-neuron translations that occur during in-place replacement accumulate hundreds or even thousands of kilometers of total discontinuous spatial translation of function, despite the possible presumption that no spatial translation is involved. Once we have already accepted spatial transfer of function on the order of hundreds of kilometers—albeit distributed over billions of neurons—what difference does a few meters across a room really make?

The spatial translation issue leaves us with two possible conclusions. We can declare that translations below some maximum permitted distance are tolerable as a true preservation of identity while farther translations are deemed preservation failures (and furthermore invoke producing another mind where at
shorter distances no other is created!), or alternatively, we can accept translations of any distance as being functionally and metaphysically equivalent. The problem is identical to that of a temporal cutoff, namely how *identity preservation* can suddenly change to *other identity*, and at what distance the cutoff resides. Additionally, the distinction between the experience of personal identity and the judgement of survival described above applies here as well. To propose a smooth transition in degrees of personal identity relative to the distance translated from the biological brain to the new substrate is to miss the point that discrete *survival* is the fundamental concern.

The rationale for a spatial transition is as problematic as for a temporal cutoff. A cutoff requires that some metaphysical property of personal identity can tolerate spatially discontinuous translation over a distance of microns, but seemingly not over a distance of meters or kilometers—or light-years (including vast accumulations of distance gathered from billions of micron-scale translations, as mentioned above). Bear in mind that the counterproposal that the distinction lies in piecewise replacement, not distance, puts us right back in the temporal issue all over again. What could possibly be the nonphysical property of consciousness and personal identity survival that is subject to these physical and spatial effects? Until solid theories are established on such questions, the default philosophical stance should be the simpler theory, namely to dispense with translation-distance-dependent identity entirely (Occam had some choice thoughts on such matters). Furthermore, doesn’t it seem remarkably suspicious that in such pondering we are tempted to assign a cutoff such that distances we can’t easily distinguish in our daily experience (microns) are considered safe while distances we casually comprehend as distinct translations (meters, kilometers, etc.) are considered unsafe? Such a designation smacks of anthropocentric reasoning. Would aliens who are a few orders of magnitude smaller or larger than humans assign this cutoff differently?

**Conclusion**

If we disregard a spatial cutoff, we conclude that QGR and SaC are functionally equivalent. If both the temporal and spatial distinctions fall, thereby equating SGR and QGR, then SGR is also functionally equivalent to SaC. This conclusion does not prove that SaC must be regarded as a successful transfer of personal identity, but it does demonstrate that both procedures should be judged in the same manner: we either grant SaC successful status and stop denigrating it as a mere copy lacking in proper identity status, or we refuse to grant identity status to the oft-favored SGR and deem both procedures to be metaphysical impossibilities.

Toward resolving this final question, the general notion of identity appears to tolerate piecewise replacement, spatial translations, and even whole parcel SaC. While the classic thought experiment of The Ship of Theseus merely poses as an open question whether identity can survive piecewise replacement, many conceptually identical examples are taken for granted to do so, such as our own bodies, in which very little matter persists over the long term, houses under lifelong renovation, waves in any physical medium, colonies, etc. With regard to SaC, multiple copies of a book or recordings of a song are casually regarded as multiple physical instantiations of a singleton information pattern (even in this very sentence we referred to the book and the song in singular vernacular, and all readers, including detractors on our central issue, accepted that phrasing without even noticing the irony). In other words, we apply the word *copy* to the physical instance of a book while recognizing that the underlying physically embedded information has neither *duplicated* nor even really *changed location* in a meaningful sense; those are properties of the physical instantiation, not the associated information. All of these examples suggest that we should not only grant the same status in all the scenarios this article has considered, but specifically a status of preserved or transferred personal identity.

One final thought on this matter involves the paradox of allowing transfer of identity in a SaC scenario that may not require destroying the biological brain and mind, a possibility this article did not discuss at length. Some readers, and most detractors on the question of transfer status in SaC scenarios,
take this paradox as proof ipso facto that identity transfer is simply beyond the pale in any SaC scenario. We propose that the best solution to the paradox of mind uploading in which the biological brain survives (including under-appreciated variants of SGR in which the brain survives [Wiley 2014]) is to adopt a completely different third model of identity that conforms to neither the transfer (preservation) nor the copy (other) interpretation, but rather to the notion that minds and personal identity can conceptually split into multiple descendants of equal primacy to their common ancestral mind and identity. A thorough description of this recommended identity model will not fit here, but please see Cerullo [2015] or Wiley [2014] for detailed descriptions.

Acknowledgements

We would like to thank Michael Cerullo, Alexander McLin, and Oge Nnadi for their collaboration on this topic and on refining and editing this paper.

References

Blackford, R., & Broderick, D. (eds.) (Aug 2014) *Intelligence Unbound: The Future of Uploaded and Machine Minds*, Wiley-Blackwell.

Blackmore, S. (2002) There is no stream of consciousness, *Journal of Consciousness Studies*, 9(5–6).

Block, N. (1995) On a confusion about a function of consciousness, *Behavioral and Brain Sciences*, 18(2), pp. 227–287.

Cerullo, M. A. (Feb 2015) Uploading and branching identity, *Minds and Machines*, Springer Netherlands, 25(1), pp. 17–36. http://link.springer.com/article/10.1007%2Fs11023-014-9352-8

Chalmers, D. (2010) The Singularity: a philosophical analysis, *Journal of Consciousness Studies*, 17(9–10), pp. 7–65.

Corabi, J. & Schneider, S. (2012) The metaphysics of uploading, *Journal of Consciousness Studies*, 19(7–8), pp. 26–44.

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

Dorrier, J. (Jan 2015) If you upload your mind to a computer—are you still you? *Singularity Hub*, Hill, D. (ed.). http://singularityhub.com/2015/01/25/if-you-upload-your-mind-to-a-computer-are-you-still-you

Hay, M. (Apr 2014) Mind uploading won’t lead to immortality, *H+ Magazine*, Rothman, P. (ed.). http://hplusmagazine.com/2014/04/24/mind-uploading-wont-lead-to-immortality

Hopkins, P. (2012) Why uploading will not work, or, the ghosts haunting transhumanism, *International Journal of Machine Consciousness*, 4(1).

Josh (2009) Uploading your mind, *Human Enhancement and Biopolitics*. https://hplusbiopolitics.wordpress.com/2009/08/08/uploading-your-min

Koene, R., & Deca, D. (Apr 2014) Whole Brain Emulation seeks to implement a mind and its general intelligence through system identification, *Journal of Artificial General Intelligence*, 4(3), pp. 1–9, ISSN (Online) 1946-0163, DOI: 10.2478/jagi-2013-0012.

Moravec, H. (1988) *Mind Children: The Future of Robot and Human Intelligence*, Harvard University Press, Cambridge.

Morris, S. (Sep 2013) Mind uploading and identity, *Singularity Weblog*, Danaylov, N. (ed.). https://www.singularityweblog.com/mind-uploading-and-identity

Olson, E. T. (Winter 2010) Personal identity, *The Stanford Encyclopedia of Philosophy*, Zalta, E. N. (ed.). http://plato.stanford.edu/archives/win2010/entries/identity-personal

Sandberg, A., & Bostrom, N. (2008) *Whole Brain Emulation: A Roadmap*, Technical Report #2008-3, Future of Humanity Institute, Oxford Univ. http://www.fhi.ox.ac.uk/brain-emulation-roadmap-report.pdf

Searle J. (1997) *The Mystery of Consciousness*, The New York Review of Books.

Schneider, S. (Mar 2014) The philosophy of ‘Her’, *The New York Times Opinionator*. http://opinionator.blogs.nytimes.com/2014/03/02/the-philosophy-of-her

Strout, J. (Aug 1997) The philosophy and technology of mind uploading, *Sci-Fi Arizona, Writer's Workshop*. http://www.scifi-az.com/articles/guest1.pdf (URL may be defunct)

Wiley, K. B. (Sep 2014) *A Taxonomy and Metaphysics of Mind-Uploading*, Humanity+ Press and Alautun Press. http://www.amazon.com/dp/0692279849

Van Gulick, R. (Spring 2014) Consciousness, *The Stanford Encyclopedia of Philosophy*, Zalta, E. N. (ed.). http://plato.stanford.edu/archives/spr2014/entries/consciousness