CHAPTER 16

COULD YOU MERGE WITH AI? REFLECTIONS ON THE SINGULARITY AND RADICAL BRAIN ENHANCEMENT

CODY TURNER AND SUSAN SCHNEIDER

In science fiction stories, such as *Star Wars* and *The Jetsons*, humans are surrounded by sophisticated AIs, but they remain unenhanced. The historian Michael Bess says these stories fall prey to a "Jetsons Fallacy"—they assume that the brain will remain the same, merely being subject to the relatively slow pace of Darwinian evolution. More realistically however, AI will not just change the world, it will likely transform the brain's cognitive and perceptual abilities as well.[1]

Consider that if we use AI technologies to transform the mind, then it will be *intelligently designed*. But we, not a god, will be the designers. So if we are to embark upon this path, we had better think it through (Schneider 2019a). The suggestion that humans should eventually merge with AI is currently discussed by researchers and the media as both as a way for humans to avoid AI-based technological unemployment and as a path to radical longevity and superintelligence. For example, Elon Musk recently remarked that humans can avoid being outmoded by AI by "having some sort of merger of biological intelligence and machine intelligence."[2] Further, he's founded a new company, Neuralink, which aims to connect the brain directly to computers. In addition, there are already many projects developing brain-implant technologies to treat mental illness,

[1] Michael Bess, *Our Grandchildren Redesigned* (Boston, MA: Beacon Press, 2015).

[2] Olivia Solon, "Elon Musk Says Humans Must Become Cyborgs to Stay Relevant. Is He Right?," *The Guardian* (February 15, 2017), https://www.theguardian.com/technology/2017/feb/15/elon-musk-cyborgs-robots-artificial-intelligence-is-he-right.

motion-based impairments, strokes, dementia, autism, and more. We are not suggesting that AI-based brain enhancements will become commonplace during the 2020's, but things may very well be moving in that direction, and the medical treatments of today will likely give rise to the enhancements of tomorrow.[3] In this chapter, we hope to clarify some of the philosophical issues at stake, and suggest a sensible path forward. We illustrate that merging oneself with AI could lead to perverse realizations of AI technology, such as the demise of the person who sought enhancement. And, in a positive vein, we offer ways to avoid this, at least within the context of one theory of the nature of personhood.

Here's how we will proceed. First, we provide background about the so-called "technological singularity" (first section) and outline some methods of cognitive and perceptual enhancement (second section). Then, in the third and fourth sections, we discuss several concerns about cognitive and perceptual enhancement. We then focus on the personal identity issue in more detail, offering a few practical suggestions in the fifth section, including certain ethical guidelines for the use of brain enhancement devices and taking a stance of "metaphysical humility" toward the metaphysics of personhood. In the sixth section, we then consider different ways *external cognitive artifacts* might *augment* personhood on the psychological theory of identity, comparing and contrasting the psychological continuity version of the theory with the narrative version. We conclude that while many external artifacts, such as lifelogs, can bolster psychological continuity, it is unclear whether this is the case with respect to narrative continuity. Finally, in the seventh section, we question whether more radical forms of enhancement, such as chips in the brain, could be constructed so as to maintain psychological continuity or narrative structure. We contend that while chips may be able to accomplish these tasks, these more invasive forms of enhancement raise philosophical complications that milder forms of enhancement lack (e.g., reduplication worries, the consciousness problem, and authenticity concerns), and we provisionally recommend on this basis that certain invasive, ("substrate replacing") enhancements be avoided in favor of biological enhancements.

The Technological Singularity

The development of AI has been driven by market forces and government and military strategic investments. Billions of dollars are pouring into constructing smart household assistants, robot supersoldiers, and supercomputers (Schneider 2019a). For example, the Japanese government has launched an initiative to have androids take care of the nation's elderly, in anticipation of a labor shortage. Further, AI is projected to outmode many human professions within the next several decades. According to a recent survey, the most-cited AI researchers expect AI to "carry out most human professions at least as

[3] Susan Schneider, *Artificial You: AI and the Future of Your Mind* (S.l.: Princeton University Press, 2019).

well as a typical human" within a 50 percent probability by 2050, and within a 90 percent probability by 2070.[4]

Given these market forces, and the strategic needs of various countries to stay abreast of the latest AI technologies, AI may soon advance to artificial general intelligence (AGI) within the next several decades. AGI is human-level intelligence that can combine insights from different topic areas and display flexibility and common sense reasoning. (Some take AGI to be the sort of system that processes information *just like* humans do, but the expression "AGI" should be understood more generally. What is essential is that the AI functions at least as well as humans in all or at least a key range of tasks, not that it achieve this by being precisely reverse-engineered from the brain.)

Superintelligent AI is a hypothetical form of AI that surpasses us in *all* domains: scientific reasoning, social intelligence, and more.[5] Ray Kurzweil, a transhumanist who is now a director of engineering at Google, writes vividly of a technological utopia in which benevolent superintelligence brings about the end of aging, disease, poverty, and resource scarcity.[6] However, even if one grants that AGI and superintelligence could be developed, this utopian scenario has been questioned by those posing the *control problem*—the problem of how humans can control a superintelligent system, given that the system is smarter than humans in all domains. The concern is that such a system may have goals that run contrary to human flourishing and that a superintelligence could lead to human extinction.[7]

Whether AI turns out to threaten the very existence of humanity or not, Kurzweil and other transhumanists contend that we are fast approaching a "technological singularity": a hypothetical point at which AI far surpasses human intelligence and can solve all sorts of problems we weren't able to solve before. The singularity, they stress, features unpredictable consequences for civilization and human nature. The idea of a singularity comes from the concept of a black hole, a "singular" object in space and time, and a place where normal laws of physics break down. In a similar vein, the technological singularity is supposed to generate runaway technological growth and massive alterations to civilization and the human mind.[8]

It is important to stress that human technological innovations may not be so rapid that they lead to a full-fledged singularity in which the world changes overnight. But the larger point still holds: as we move further into the twenty-first century, unenhanced humans may not be the most intelligent beings on the planet for that much longer. The greatest intelligences on the planet may be synthetic.

[4] Vincent C. Müller, and Nick Bostrom, "Future Progress in Artificial Intelligence: A Survey of Expert Opinion," *Fundamental Issues of Artificial Intelligence* (2016): 555–572.

[5] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (New York: Oxford University Press, 2014).

[6] Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (New York, NY: Viking, 2006).

[7] Bostrom, *Superintelligence*, 2014.

[8] Vernor Vinge, "The Coming Technological Singularity: How to Survive in the Post-human Era," *Whole Earth Review*, 1993.

Cognitive and Perceptual Enhancement: Some Background

Cognitive and perceptual enhancements amplify or extend one's cognitive or perceptual capacities through improvement or augmentation of one's information processing systems, including sensory systems.[9] Whereas therapies intervene to correct a problem with a cognitive or perceptual system/subsystem, enhancements, by contrast, intervene to improve a cognitive or perceptual ability, and perhaps even provide a new capacity[10]

There are many kinds of cognitive and perceptual enhancement technologies that could be utilized in the future, ranging from the ordinary to the science fiction-like. Different methods of enhancement can be summarized as follows:

1. *Brain implants involving AI technologies*. Currently, brain chips are primarily being developed for therapeutic (as opposed to enhancement) purposes. Theodore Berger's lab at the University of Southern California, for example, is developing an artificial hippocampus that could allow individuals with severe memory impairment to formulate new memories. Researchers are currently at work creating brain chips for other impairments as well, such as depression, post-traumatic stress disorder, and Alzheimer's disease. As neural prosthetic technology develops, it is likely that such technologies will be used for enhancement as well. People will wish to enhance their reasoning capacities, memory, and attention well beyond what is considered to be biologically normal.
2. *Pharmaceutical drugs*. While most pharmaceutical drugs are currently developed for therapeutic purposes (e.g., to treat ADHD), this will not in all likelihood remain the case. Certain pharmaceutical drugs are currently being used off label for enhancement purposes, such as metformin, for life extension and Adderall, for attentional enhancement. In the future, more and more drugs may be produced to enhance the brains and bodies of normal individuals.
3. *External cognitive artifacts*. These are extra-cranial devices that function to enhance human cognition. This includes numerous different technologies, such as the internet, navigation systems, cell phones, diaries, and brain-computer interface devices.
4. *Biological enhancements*. Biological enhancements can involve the use of biotechnology, including nanotechnology and genetics, to extend the lifespan of the biological brain or to augment certain parts of the brain, or alter genes of subsequent generations so parents can produce smarter offspring.

[9] Nick Bostrom, and Anders Sandberg, "Cognitive Enhancement: Methods, Ethics, Regulatory Challenges," *Science and Engineering Ethics* 15, no. 3 (2009): 311–341.

[10] The distinction between a therapy and an enhancement is controversial, and some reject it altogether, claiming that it is often difficult to discern whether a case is a therapy or an enhancement (see Bostrom & Roache, 2007).

5. Other, more commonplace, *Conventional enhancements* (e.g., education and psychological interventions). The term "enhancement" could be used broadly, including mental strategies that enhance core mental capacities. Bostrom and Sandberg observe: "The spectrum of cognitive enhancements includes not only medical interventions, but also psychological interventions (such as learned 'tricks' or mental strategies), as well as improvements of external technological and institutional structures that support cognition."[11]
6. *Mind-uploading*. A hypothetical, (and highly speculative) type of enhancement that is discussed by transhumanists, which involves the migration of a mind from a brain to a computer. Proponents of this procedure believe that the mind can be implemented onto different substrate, just as computer software programs can be implemented onto different hardware. The ultimate goal behind mind-uploading is to either to allow the mind to live in a virtual reality world or reside in a computer that operates inside (or connected to) a humanoid robot or a biological body.[12]

It is important to bear in mind that no one can accurately predict the future of brain enhancement technologies, although it is perhaps possible to make some reasonable approximations from looking at present trends and research. We are not suggesting that human brain-uploading will be developed, or even that those wishing for brain enhancements will do so through invasive AI-based techniques, rather than biological or genetic enhancements or noninvasive AI-based technologies. Bearing in mind these qualifications, in what follows, we focus on more radical and hypothetical forms of AI-based brain enhancement that may arise in or around a singularity, if such indeed occurs.

Suppose it is 2045, and you stroll into a new medical enhancement center called "The Center for Mind Design." There customers can choose from a variety of brain enhancements. Human Calculator can provide you with savant-level mathematical abilities; Zen Garden can give you the meditative states of a Zen master, and so on. It is also rumored that if clinical trials go as planned, customers may soon be able to purchase an enhancement bundle called "Merge": a series of brain enhancements allowing customers to gradually augment and transfer all of their mental functions to the cloud over a period of five years.[13]

Should you add one or more chips to your brain, and even try Merge? In the following we discuss some considerations that are relevant to your decision.

Concerns

Even assuming these enhancements are medically safe, it doesn't follow that they are beneficial to an individual or society. For instance, enhancements may only be available

[11] Bostrom and Sandberg, "Cognitive Enhancement: Methods, Ethics, Regulatory Challenges" (2009), 312.

[12] David J. Chalmers, "The Singularity: A Philosophical Analysis," *Journal of Consciousness Studies* 17 (2010): 9–10.

[13] Schneider, *Artificial You*.

for the wealthiest members of society, creating a rich-poor intellectual gap, or perhaps, in the vein of a science fiction dystopia, socially mandated microchips become the norm, so that schools, governments, or employers require certain enhancements, and even use them to mine data and track people.

These scenarios raise the concern that enhancements will dehumanize us. Indeed, authors in the cyberpunk genre of science fiction depict technological dystopias in which individuals lose control of their enhancements—governments or corporations hack their thoughts, cut off their access to their implants, and threaten their very survival.[14] This is clearly dehumanizing, and it is not hard to foresee that such technologies could lead to abuse in the hands of an authoritarian dictatorship or unregulated capitalist economy. In a different vein, one might worry that even if such scenarios are avoided, radical brain enhancements would rob us of our humanity because our very limitations and vulnerabilities are part of what makes us human in the first place. Such limitations and vulnerabilities might, for instance, preserve certain traits that ought to be preserved, like humility.[15] Relatedly, Daniel Callahan, a so-called "life cycle traditionalist," criticizes any attempts to extend the human lifespan or control the aging process via enhancement.[16]

This "traditionalist" attitude is antithetical to the aspirations of transhumanists, such as the biological gerontologist Aubrey de Grey, who views aging as a disease that we may be able to overcome in our lifetime with advances in medical technology.[17] Transhumanists, like Nick Bostrom, Anders Sandberg, James Hughes and Aubrey de Grey, claim that the human species is now in a comparatively early phase and that its very evolution will be altered by developing technologies. Future humans will have radically advanced intelligence, extreme longevity, deep friendships with AI creatures, and elective body and mental characteristics. Transhumanists share the belief that such an outcome is very desirable, both from the vantage point of one's own personal development and for the development of our species as a whole.[18] Perhaps some, like Callahan, would not wish for longevity or advanced intelligence, but transhumanists have always stressed that enhancements should be optional, and stressing the import of human flourishing, they would clearly view cyberpunk dystopias as undesirable.

Schneider agrees with many of the transhumanist aims but has doubts about whether the radical AI-based enhancements they advocate will accomplish the transhumanists goals of longevity, human flourishing, and intelligence enhancement. Her concern is that even if the technologies are medically safe and are not used as tools by surveillance

[14] William, Gibson. *Neuromancer* (New York: Ace Books, 1984).

[15] Kevin Fitzgerald, S.J., "Medical Enhancement: A Destination of Technological, Not Human, Betterment," *Medical Enhancement and Posthumanity: The International Library of Ethics, Law and Technology* (2008): 39–53.

[16] D. Callahan, "Aging and the Life Cycle: A Moral Norm?," *A World Growing Old: The Coming Health Care Challenges* (Washington: Georgetown University Press 1995), 21–27.

[17] Aubrey de Grey, *Ending Aging: the Rejuvenation Breakthroughs That Could Reverse Human Aging in Our Lifetime* (St. Martin's Griffin, 2008).

[18] The basic tenets of Transhumanism were first formally put forth by the World Transhumanist Association in the Transhumanist Declaration in 1998.

capitalism or an authoritarian dictatorship, these enhancements may still fail to do their job for philosophical reasons. In what follows, we explore one such concern, a problem that involves the nature of the self.

Personal Identity and Radical Enhancement

Imagine that, longing for superintelligence, you consider buying Merge at the Center for Mind Design. Should you do it? To understand whether you should embark upon this journey, you must first understand what and who you are. But what is a self or person? What allows a self to continue existing over time? Like consciousness, the nature of the self is a matter of intense philosophical controversy. And given your conception of a self or person, would you continue to exist after adding Merge—or would you have ceased to exist, having been replaced by someone else? If the latter, why should you try Merge in the first place?

Even if your hypothetical merger with AI brings benefits like superhuman intelligence and radical life extension, it must not involve the elimination of any of what philosophers call "essential properties"—the things that make you.[19] Even if you would like to become superintelligent, knowingly trading away one or more of your essential properties would be tantamount to suicide—that is, to your intentionally causing yourself to cease to exist. So before you attempt to redesign your mind, you'd better know what your essential properties are.

So what are your essential properties? Unfortunately, there is intense disagreement on the matter. One can distinguish between at least four influential approaches to personal identity in the metaphysics literature:

Brain-based materialism: You are essentially the material that you are made out of (i.e., your body and brain).[20, 21]

Dualist theories: Views that explain personal identity in terms of the persistence of an immaterial or nonphysical substance (such as a soul or Cartesian ego).[22]

Psychological theories: Views that explain personal identity in terms of psychological properties, such as experiences, beliefs, memories, and so forth.[23]

[19] Joseph Corabi and Susan Schneider, "Metaphysics of Uploading," *Journal of Consciousness Studies* 19 (2012): 26.

[20] A.J. Ayer, *Language, Truth, and Logic* (London: Gollancz, 1936).

[21] J.J. Thomson, "People and Their Bodies," *Reading Parfit* (ed. J. Dancy, Oxford: Blackwell, 1997).

[22] R.G. Swinburne, "Personal identity," *Proceedings of the Aristotelian Society* 74 (1973): 231–247.

[23] John Locke, *An Essay Concerning Human Understanding* (ed. P.H. Nidditch, 4th ed., Oxford: Clarendon Press, 1975).

The No Self View: The self is an illusion. The "I" is a grammatical fiction (Nietzsche). There are bundles of impressions, but there is no underlying self (Hume). There is no survival because there is no person (Buddha).[24]

Each of these positions has its own implications about whether to enhance the brain. For example, suppose you are partial to the soul theory. In this case, your decision to enhance would seem to depend on whether you have justification for believing that your enhanced brain and body would retain your soul or immaterial mind.

Many philosophers sympathize with the "psychological continuity view," which is one type of psychological theory. We will discuss psychological theories in more detail shortly. But for now, the psychological continuity view says that the holding of a certain psychological relation is necessary or sufficient, or both, for an individual to persist over time—you survive by inheriting mental features such as memories, beliefs, personality dispositions and so on.[25] But this means that if we change our memories or personality in radical ways by enhancing the brain, the continuity could be broken.

Alternately, consider brain-based materialism. Within the fields of philosophy of mind and metaphysics, views that are materialist claim that minds are basically physical or material in nature and that mental features, such as the thought that Bach is a famous composer, are ultimately just physical features. (This position is often called "physicalism" as well.) Brain-based materialism says this, and, in addition, it makes the additional claim that your thinking is dependent on the brain. Thought doesn't "transfer" to a different substrate. So on this view, enhancements should not change one's material substrate, or the person would cease to exist. So enhancements like Merge are unsafe, because you are replacing parts of your brain with AI components.

Advocates of a mind-machine merger tend to reject the view that the mind is the brain, however. They believe that the mind is like a software program: just as you can upload and download a computer file, your mind can add new lines of code and even be uploaded onto the cloud. According to this view, the underlying substrate that runs your "self program" doesn't really matter—it could be a biological brain or a silicon computer.

However, we believe that this computationalist view of the mind doesn't hold up under scrutiny. A program is a list of instructions in a programming language that tell the computer what tasks to do, and a line of code is like a mathematical equation. It is highly abstract, in contrast with the concrete physical world. Equations and programs are what philosophers call "abstract entities"—things not situated in space or time. But minds and selves are spatial beings and causal agents; our minds have thoughts that cause us to act in the concrete world. And moments pass for us—we are temporal beings.[26]

[24] For a transhumanist approach, see Hughes, James, "Humanism for Personhood: Against Human-Racism," *Free Inquiry* 24 (2004).

[25] D. Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984).

[26] Schneider, *Artificial You*.

Perhaps you are inclined to the No Self View. In this case, survival isn't an issue for you, and you can make enhancement decisions solely based on other considerations, such as maximizing the happiness of future sentient beings and minimizing suffering.

So, how can we approach the issue, given all this philosophical disagreement? Would you survive Zen Garden? Merge? You might feel inclined to passionately defend a certain theory of personal identity if you chat with your friends, colleagues or students about these issues, but would you put your money where your mouth is?

Suggestions

We have three suggestions.

1. In Making Radical Brain Enhancement Decisions, Distinguish the Issue of Personal Identity, or Survival over Time, from that of Consciousness

Notice that the question of whether or not your identity survives cognitive enhancement—whether that future being is really *you*—is distinct from the question of whether or not consciousness survives. It is currently unclear whether AI can be conscious. If it is, then microchips can, at least in principle, be used in areas of the brain responsible for consciousness without one losing consciousness or experiencing diminished consciousness. It is possible that attempts at radical enhancement, such as mind-uploading or the augmentation of many of one's mental abilities through implantation of AI devices, that consciousness is preserved, but personal identity is not. Perhaps the uploaded copy of your mind is conscious, but the copy is still not you.

Schneider believes it will be easier to tell if AI is conscious than it will be to determine which theory of personal identity is true, if any. This is because she suspects we can test whether consciousness could have a different substrate. Schneider has devised a test for synthetic consciousness, which she calls "the chip test."[27] The test involves observing whether normal patients having AI components placed in their brains (in place of neural tissue, which is removed) experience a loss of consciousness after the surgery: "If . . . a prosthetic part of the brain ceases to function normally—specifically, if it ceases to give rise to the aspect of consciousness that that brain area is responsible for—then, there should be behavioral indications, including verbal reports . . . This would indicate a 'substitution failure' of the artificial part for the original component. Microchips of that sort just don't seem to be the right stuff."[28] Similarly, patients needing prosthetic devices in

[27] Schneider, *Artificial You*. [28] *Ibid*, 54–55.

parts of the brain responsible for consciousness to correct a problem due to brain injury or disease may experience a restoration of elements of their conscious experience. Like the episodes Oliver Sacks wrote about, patients can report changes to their consciousness, and they can be carefully tested by researchers to mark alterations in conscious brain processing.

In contrast, it is difficult to envision testing different theories of personal identity. After all, we cannot expect behavioral differences between a person and her conscious upload, molecular duplicate, functional isomorph, and so on. Such will likely believe they are the same person they were before, as they have all the same memories and behavioral traits. Instead, we have to rely on armchair philosophical considerations to adjudicate between competing theories. But the problem of personal identity has been intensely debated by philosophers for centuries, and it has proven to be vexing, as we have seen, and there is intense disagreement over the different solutions. In light of this we suggest the following approach.

2. A Stance of *Metaphysical Humility*

In *Artificial You*, Schneider opts for a stance of "metaphysical humility" in the face of radical brain enhancements. Given the controversies over personal identity, claims about survival that involve one "transferring" one's mind to a new type of substrate or making drastic alterations to one's brain must be carefully scrutinized. As alluring as greatly enhanced intelligence or digital immortality may be, there is simply too much disagreement in the personal-identity literature over whether any of these "enhancements" would extend life or terminate it.

All this uncertainty suggests that one should take the transhumanist approach to radical enhancement with a grain of salt. Enhancements like brain-uploading or adding brain chips to augment intelligence or one's perceptual abilities are key enhancements invoked by the transhumanists, yet these enhancements sound strangely like the thought experiments philosophers have used for years as problem cases for various theories of the nature of persons. In light of this, it isn't surprising to us that the enhancements aren't as attractive as they might seem at first.[29]

The way forward is public dialogue, informed by metaphysical theorizing as well as a technical understanding of AI/neurotechnologies. This may sound like a sort of intellectual cop-out, like we are throwing our hands up in the face of ignorance, but we are not saying that further metaphysical theorizing is useless. To the contrary, we believe the first step is to underscore the life-and-death import of further metaphysical reflection on these issues: ordinary individuals must be capable of making informed decisions about enhancement. If the success of an enhancement rests on (inter alia) classic philosophical issues that are difficult to solve, the public needs to realize this, and not assume that researchers, members of the media or business leaders who are enthused by the

[29] *Ibid.*

bells and whistles of a new technology are also experts on philosophical questions of whether one should enhance.

3. Support Regulations of Brain Enhancement Devices that Require that Consumers Be Informed about the Personal Identity Debate

Bearing this in mind, brain-enhancement devices should be regulated by a government agency, such as the Food and Drug Administration in the United States, and disclosure of the personal identity controversy should be required, just as medical risks for pharmaceutical drugs are required to be disclosed. Consider, for instance, that patients routinely grapple with ethical issues when they consider whether to undergo genetic testing, asking themselves whether they or a loved one would really want to know if they were going to have a high probability of getting a certain horrible illness, what to do if life insurance companies get hold of their data, and so on. For this reason, it is protocol at many medical centers in the United States that patients considering genetic testing be required to meet with a genetics counselor or nurse who discusses the pros and cons of testing before testing and then return and meet with the counselor to discuss the test results. In the context of brain-enhancement devices, we believe a similar approach could be taken.

We have further suggestions as well. But for now, let's assume that you are inclined to resist our suggestion of metaphysical humility: in particular, you are strongly persuaded by the psychological view. If so, we have further suggestions for you.

A Way Forward? The Psychological Continuity and Narrative Views

Suppose that, in addition to being impressed by the psychological view, you've just learned that individuals using AI-based enhancements are doing so without a loss of conscious experience. On the assumption that a certain version of the psychological view obtains, perhaps certain kinds of brain enhancements could *enhance* psychological continuity, reducing the likelihood that numerical identity would not obtain after the enhancement.

To see what we have in mind, we will need to distinguish different versions of the psychological theory. There are two main versions: psychological continuity views and narrative views. We've already introduced continuity views, in broad strokes. Psychological continuity views differ with respect to which direct connection is the most important in terms of constituting personal identity. While all psychological continuity theorists

believe that the connection of memory is necessary for personal identity, some go so far as to claim that memory is the only relevant psychological connection when it comes to personal identity.

Psychological continuity views of identity can be contrasted with *narrative views*. Narrative views concur that the relationship of psychological connectedness is necessary for personal identity but deny that it is sufficient. Proponents of a narrative view hold that personal identity additionally requires the relationship of narrative connectedness. Two of the most prominent defenders of the narrative view are Marya Schechtman and Anthony Rudd. Both Schechtman and Rudd hold that narrative connectedness exists when one is equipped with an integrative story about themselves which details the chronology of their lives and highlights the most important memories/time slices contained within that chronology. Rudd analogizes this "integrative story" to a Cartesian ego. The idea is not that narratives are metaphysically immaterial entities in the same way that Cartesian egos are, but simply that narratives *function* like Cartesian egos by providing us with a unified sense of personhood.[30] Schechtman, on the other hand, views the narrative as an extended story which transcends the scope of any particular subset of time slices. Schechtman writes: "It is by no means obvious that the most essential part of a person's experience at any time can be reproduced in an independent time-slice, even if we imagine that slice containing all of the relevant forward- and backward-looking elements. . . . [Our experience] is essentially something that takes place over time, and whose relevant attributes cannot be caught in a moment or even a series of moments."[31]

The main difference between the narrative theory and the psychological continuity theory is that the former views personhood as more active and self-constructed than the latter. Psychological continuity theories see personhood as a fundamentally passive phenomenon that is constituted by relations of psychological connectedness. Subjects are not responsible for establishing the relevant relations of psychological connectedness through the creation of a narrative. Narrative views, on the other hand, claim that subjects are able to actively interpret and construct their own identities by choosing which narrative explanation best suits their life.

Bearing in mind these two versions of the view, we will explore how, should a psychological theory be correct, various memory enhancing external cognitive artifacts may function to undermine, preserve, or bolster personhood. We begin with the sort of artifacts around us now, and then apply the points we make to the case of radical brain enhancements. There are currently many different kinds of external artifacts which function to enhance memory, including the internet, navigation systems, cell phones, diaries, and brain-computer interface devices. We will first consider how memory enhancing external artifacts may undermine personhood (again, we assume the psychological theory of personhood) before suggesting how this may be countered. More specifically: we argue that personhood is at a greater risk of being undermined by

[30] Anthony Rudd, "In Defence of Narrative," *European Journal of Philosophy* 17 (2009): 60–75.
[31] Marya Schechtman, *The Constitution of Selves* (Ithaca, NY: Cornell University Press, 1996).

memory enhancing external artifacts on the narrative view than it is on the psychological continuity view. Then, we illustrate how a particular memory enhancing external artifact, the visual lifelog, bolsters personhood if (a) the memories that are stored in visual lifelogs are nonrepresentational, but (b) the memories stored in biological memory are representational.

Nicolas Carr contends that such artifacts weaken personhood by making us less intellectually autonomous: "When we outsource our memory to a machine, we also outsource a very important part of our intellect and even our identity."[32] Intellectual autonomy, broadly speaking, is the ability to think for oneself and to not be overly reliant on other people and external devices when formulating beliefs and engaging in cognition.[33] The main way in which memory enhancing external artifacts make us less autonomous, according to Carr, is by rendering us less knowledgeable. The internet, in particular, makes us less knowledgeable by minimizing the amount of information that we need to store in biological memory.[34]

However, even if Carr is correct, while intellectual autonomy and personhood are related, they do not necessarily go hand in hand. More specifically, if the psychological continuity theory is assumed, then personhood may be boosted by memory enhancing artifacts, *even if* Carr is correct that these artifacts undermine intellectual autonomy. Recall that personhood, according to the psychological continuity view, is explained in terms of psychological connectedness. Memory enhancing external artifacts such as the internet and iPhones could strengthen relations of psychological connectedness by allowing subjects to unearth memories that would have otherwise been forgotten. Again, this holds true despite the fact that the artifacts may simultaneously function to undermine intellectual autonomy in various ways. Consider, for example, an Alzheimer's patient who is gradually losing her biological memory. Such a patient might use an external artifact to help her preserve psychological continuity. This is indeed the situation depicted in Clark and Chalmers' fictional case of Otto and Inga, which they use not as an example of how personhood can be preserved by enhancements but as an argument for the extended mind hypothesis.[35]

Further, it isn't clear that autonomy is really undermined in these cases. This seems to depend on deep issues about whether the mind could be extended. To see what we have in mind, consider the Alzheimer's patient case. Is the autonomy of someone who is losing their memories really undermined here? In a sense, it seems not, at least in one sense of "autonomy," as the technology preserves their independence. Still, it is correct that the person is not autonomous in another sense, as they are now dependent on an external

[32] Nicholas G. Carr, *The Shallows: What the Internet Is Doing to Our Brains* (W.W. Norton, 2011), 9.

[33] See also Michael P. Lynch. *The Internet of us: knowing more and understanding less in the age of big data* (New York: Liveright Publishing Corporation, a division of W.W. Norton & Company, 2016).

[34] Contra Carr, we believe that the internet increases the knowledge at our fingertips, as we can look anything up on the web, and we can still remember our results. In any case, Carr's idea is that we become more reliant on external artifacts as these artifacts become increasingly integrated into our cognitive lives.

[35] Andy Clark, and David J. Chalmers, "The Extended Mind," *Analysis* 58 (1998): 7–19.

device for cognition. How would we decide whether there is an overall loss of autonomy in such cases? It seems that if the external device is an extension of the patient's cognition, then the device arguably makes the patient more autonomous. In that case, the person isn't dependent on an external device because the enhancement is actually part of their own cognitive system.

In addition to helping subjects unearth memories that would have otherwise been forgotten, external artifacts can also give subjects access to digital memories that are more fine-grained than those stored in biological memory. Digital memories (like those stored on Facebook) are photographic images, and photographic images are arguably more than just mere representations of previous perceptions. Kendall Walton (1984) argues for what he calls "photographic realism," which holds that a photographic image of X allows one to indirectly see X itself (as opposed to directly see a representation of X): "We *really do, literally*, see our deceased ancestors when we see photographs of them."[36] Walton argues for photographic realism on the basis of providing a conceptual analysis of what it means to have "perceptual contact" with the world. If he is correct, then the digital memories stored in external artifacts are not mere representations of past perceptions, but are rather re-presentations or "fixed reflections" of those perceptions. Biological memories, by contrast, are in all likelihood representations of previous events. This position is supported by the causal theory of memory, which is the default view of memory in contemporary philosophy.[37] According to the causal theory, remembering requires a causal connection between the original experience remembered and the consequent representation of that experience in memory. It is worth pointing out, to be fair, that not all theories of memory take memories to be representational by nature. The empiricist theory, for example, contends that memories are "preserved sense impressions."[38] Mohan Matthen, however, argues against the idea that memories are "preserved content" by emphasizing that a single biological memory can occur in a myriad of different formats.[39] If it is true that (a) digital memories are transparent in the sense advocated by Walton, and (b) biological memories are representational, then it is arguably the case that the former kind of memory is more "real" than the latter. One could contend, in particular, very much in the vein of Plato's concept of "mimesis," that representations are always less real than the items represented. Of course, videos can be altered and edited, as has been increasingly seen in the so-called "fake news" era. This does not undermine the argument that unaltered videos are more transparent than biological memories though. All in all, this argument serves to lend further support

[36] Kendall L. Walton, "Transparent Pictures: On the Nature of Photographic Realism," *Noûs* 18 (1984): 67–72.

[37] See C.B. Martin, and Max Deutscher, "Remembering," *Philosophical Review* 75 (April 1966): 161–196; Sven Bernecker, *Memory: A Philosophical Study* (Oxford University Press, 2010).

[38] David Hume & D.G.C. Macnabb (eds.) *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning Into Moral Subjects* (Collins, 1739).

[39] Mohan Matthen, "Is Memory Preservation?," *Philosophical Studies* 148(1) (2010): 3–14.

to the hypothesis that external artifacts can bolster personhood on the psychological continuity view.

Things may be different when it comes to the narrative view of identity. The narrative view, to reiterate, explains personhood primarily in terms of narrative connectedness. While intellectual autonomy is conceptually distinct from psychological connectedness, it may not be fully conceptually distinct from narrative connectedness. This is because narrative connectedness requires active cognitive interpretation and construction on the part of the subject. Or, to put it differently, narrative connectedness appears to involve the execution of intellectually autonomous acts. It stands to reason, then, that by undermining intellectual autonomy, certain memory enhancing external artifacts may also undermine personhood on the narrative view.

Here, it is helpful to consider a particular memory enhancing external artifact: lifelogs. Lifelogs are devices that record one's personal experiences from the first person point of view. There are various different kinds of such devices: "A key example is SenseCam, a small wide-angle camera worn around one's neck, taking a picture with a certain interval or when its sensor detects some environmental change. These pictures are then edited into a visual lifelong with certain narrative structure, transforming, aiding, and in some cases constituting one's autobiographical narrative."[40] Lifelogs are unique in that they serve as external aids to both biological memory and narrative structure. In other words, lifelogs develop a narrative explanation of one's memories for the subject. Certain social media sites, such as Facebook, already accomplish this task to some extent by integrating one's pictures together to form a story. The increasing integration of lifelogs and related technologies into our lives may lead subjects to become more dependent on artifacts for their personal narratives, for better or worse. After all, if artifacts are crafting narrative explanations for subjects, then there may be less of a need, or at least less motivation, for subjects to craft their own narrative explanations. In this case, narrative explanations would become biographical as opposed to autobiographical.

The partial offloading of narrative structure to external devices certainly undercuts intellectual autonomy; the question is whether it undercuts narrative connectedness as well. If this offloading procedure does undermine narrative connectedness, then it also undermines personhood on the narrative view of identity. One might deny, however, that narrative connectedness necessitates intellectual autonomy. Perhaps partially offloading narrative structure to external artifacts can strengthen narrative connectedness in a similar way that partially offloading biological memory can strengthen psychological connectedness. Recall that narrative connectedness exists when a subject is able to provide a narrative explanation of the chronology of their lives and experiences. One might argue that external artifacts can assist subjects in providing this narrative explanation and that it does not matter whether or not the subject is personally responsible for constructing the narrative explanation.

[40] Richard Heersmink, "Distributed Selves: Personal Identity and Extended Memory Systems," *Synthese* 194, no. 8 (2017): 3136.

Caveats

Now let us ask: could the enhancements of the future, such as brain chips, be constructed to maintain continuity or narrative structure? If the psychological theory of personal identity is correct, and if technologies like brain chips can be made to preserve psychological properties like memories and personality traits, then it seems as if more radical enhancements also have the potential to preserve/bolster personal identity in the manner described in the previous section. It may even be possible to design a chip that preserves narrative structure.

We must proceed carefully though. First, it is not clear if chips would preserve consciousness, when used in parts of the brain that are part of the neural basis of conscious experience. If someone replaces these parts, important psychological properties (experiential properties) would be lacking. It would be dubious to see the future zombie as a person or having a mind, let alone the same person as before. Second, we've indicated that psychological views are controversial. In particular, they face "reduplication problems"—problems involving thought experiments in which one's pattern, narrative or psychological configuration is copied so precisely that, by the light of the psychological views, there seems to be two or more instances of the same individual at the same time.[41]

Third, brain chips and other more radical forms of enhancement may raise concerns related to authenticity that milder forms of enhancement lack. Imagine a brain chip that enables you to not only unearth memories that would have otherwise been forgotten but also consciously access many more memories over a given time interval than you would have been able to without the chip. One concern about such a chip is that it may incentivize people to not be mindful and to instead "live in the past." Insofar as authenticity is connected with mindfulness (as existentialists like Sartre claim), such a chip will function to make people less authentic. This worry, to be sure, also exists in the case of external artifacts, but is magnified in the case of brain chips that directly affect cognition.

Another "authenticity" related worry concerns the possibility that radical enhancements will augment psychological suffering. While neural prosthetics which raise our IQ levels or make us faster thinkers have obvious benefits, they may also function to amplify the "cognitive noise" which is responsible for the majority of psychological suffering within our species. Put differently, if the Buddhists are on the right track in claiming that all suffering is born out of thinking, then it is plausible that making us faster or better thinkers via brain chips will increase psychological suffering by and large (as opposed to leading to enlightenment and wisdom). Of course, particular kinds of brain chips, like the Zen Garden chip mentioned previously, might be immune to these worries concerning mindfulness and suffering.

Fourth, consider that, from the vantage point of the brain view, if you have these chips, and they replace parts of the biological brain, there will be a point at which the biological brain is so diminished that instead of ensuring continuity over time, you

[41] See Parfit 1984, Sider 2001, Olson 2007, and Schneider 2019a.

would inadvertently end your life. Bearing in mind our stance of metaphysical humility, it would be unwise to rule out the possibility that the mind is the brain, for the brain is responsible for human cognitive and perceptual processing, making this position quite plausible. This leads us to suggest the following.

Don't Offload Parts of the Biological Brain, Insofar as You Suspect that the Brain View May Be Correct

Even if AI is capable of underlying conscious experience, AI-based enhancements, if used, should supplement the workings of intact brain tissue, not destroy it and offload its activities to the cloud or another AI device. Biological therapies could instead be utilized to extend the life of the biological brain, or AI components could supplement activities of the brain, without replacing tissue. (Bearing in mind the earlier caveat that too radical of enhancements of these latter sorts may still be incompatible with survival over time, depending upon what one's essential properties are.)

Conclusion: A Humble Approach

It would be optimal if we could provide you with a clear, uncontroversial path to guide you through the brain enhancement decisions. Instead our message today has been: As we consider enhancement decisions, we must do so, first and foremost, with a mindset of metaphysical humility. Remember how controversial the different theories of personal identity are.

Still, we have offered several provisional recommendations. We proposed that in making enhancement decisions, it is important to distinguish the issue of personal identity from that of consciousness. We also suggested that future consumers considering such enhancements be educated about the personal identity debates, as well as medical risks. In addition, we outlined various ways in which enhancements may be capable of preserving personhood if a psychological view is correct. Enhancements, in particular, may be able to strengthen relations of psychological continuity and perhaps even narrative structure. This assumes the controversial view that a psychological theory of personal identity is correct, however. Further, if the brain theory is correct, these enhancements may be problematic, if they involve replacing parts of the brain. In light of this, and bearing in mind the discussion of metaphysical humble approach, we believe it is most sensible that future enhancements *both* preserve continuity while not replacing parts of the brain may be safest.

Bibliography

Bess, Michael. *Our Grandchildren Redesigned*. Boston: Beacon Press, 2015.

Bostrom, Nick ed. *Superintelligence: Paths, Dangers, Strategies*. New York: Oxford University Press, 2014.

Bostrom, Nick and Rebecca Roache. "Ethical Issues in Human Enhancement." In *New Waves in Applied Ethics*, edited by J. Ryberg, T. Petersen, and C. Wolf, 120–52. London: Palgrave-Macmillan, 2007.

Bostrom, Nick and Anders Sandberg. "Smart Policy: Cognitive Enhancement and the Public Interest." In *Enhancing Human Capabilities*, ed. Julian Savulescu, Ruud ter Muelen, and Guy Kahane. Hoboken: Wiley-Blackwell, 2009.

Carr, Nicholas G. *The Shallows: What the Internet Is Doing to Our Brains*. W.W. Norton, 2011.

Chalmers, David J. "Facing up to the Problem of Consciousness." *Journal of Consciousness Studies* 2(3) (1995): 200–19.

Chalmers, David J. "The Singularity: A Philosophical Analysis." *Journal of Consciousness Studies* 17 (9–10) (2010): 9–10.

Clark, Andy and Chalmers, David J. "The Extended Mind." *Analysis* 58 (1) (1998): 7–19.

de Grey, Aubrey. *Ending Aging: the Rejuvenation Breakthroughs That Could Reverse Human Aging in Our Lifetime*. New York: St. Martin's Griffin, 2008.

Fitzgerald, K. "Medical Enhancement: A Destination of Technological, Not Human, Betterment." In *Medical Enhancement and Post-Modernity*, ed. B. Gordijn and R. Chadwick, 39–55. Dordrecht: Springer, 2008.

Gibson, William. *Neuromancer*. New York: Ace Books, 1984.

Heersmink, Richard. "Distributed Selves: Personal Identity and Extended Memory Systems." *Synthese* 194.8 (2017): 3135–51.

Hume, David & D.G.C. Macnabb, eds. *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning Into Moral Subjects*. Glasgow: Collins, 1739.

Kurzweil, Ray. *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking, 2005.

Locke, John. *An Essay concerning Human Understanding*, 4th ed., ed. P.H. Nidditch. Oxford: Clarendon Press, 1975.

Lynch, Michael. *The Internet of us: knowing more and understanding less in the age of big data*. New York: Liveright Publishing Corporation, a division of W.W. Norton & Company, 2016.

Olson, Eric T. *What Are We?: A Study in Personal Ontology*. Oxford University Press, 2007.

Parfit, D. *Reasons and Persons*. Oxford: Clarendon Press, 1984.

Rucker, Rudy. *Software*. New York: Eos, 2001.

Rudd, Anthony. "In Defence of Narrative." *European Journal of Philosophy* 17(1) (2009): 60–75.

Schechtman, Marya. "Stories, Lives, and Basic Survival: A Refinement and Defense of the Narrative View." *Royal Institute of Philosophy Supplement* 60 (2007): 155–78.

Schneider, Susan. *Artificial You: AI and the Future of Your Mind*. Princeton: Princeton University Press, 2019a.

Schneider, Susan. "Can You Add a Microchip to Your Brain?" *New York Times*, June 2019b.

Sider, Theodore. "Criteria of Personal Identity and the Limits of Conceptual Analysis." *Philosophical Perspectives* 15(s15) (2001): 189–209.

Swinburne, R.G. "Personal Identity." *Proceedings of the Aristotelian Society* 74 (1973): 231–47.

Vinge, Vernor. "The Coming Technological Singularity: How to Survive in the Post-human Era." *Whole Earth Review*, 1993.

Walton, Kendall L. "Transparent Pictures: On the Nature of Photographic Realism." *Noûs* 18(1) (1984): 67–72.