

# Metaphysical Daring as a Posthuman Survival Strategy

PETE MANDIK

## 1. BACK UP TO LIVE AND LIVE TO BACK UP

Congratulations! You have survived reading all of the way to the end of this sentence. Along the way you have endured many changes: slight changes in your memory and experience, tiny fluctuations in your temperature and mass, minor alterations in the number and distribution of molecules in your body. Like me, you have likely survived even more significant changes. Birth was a big one, though I don't myself remember my own. However, I do recall a very bad appendicitis and ensuing appendectomy that I survived in my early 20s. I survived the removal and disposal of one of my organs. If you deny that people frequently survive such changes, you are likely employing a bizarre and overly philosophical notion of personal survival. But even with an ordinary sense of survival in view, we can raise this question: How much change *can* a human undergo and still survive? Of particular interest in this article is a question concerning a very dramatic bodily change: Are there any circumstances under which you could survive the removal and disposal of your brain?

One of the many ideas to make the journey from the pen of a science fiction author to the page of a philosophy journal is the idea of mind-uploading, a hypothetical means of converting a human mind from a biological substrate to a computational one, with a promise of extending one's life indefinitely through frequent backups. That a human might survive the death of the body is an ancient idea, enshrined in many religious and spiritual traditions. The secular and computational version of the idea of life after death—the very idea of mind “uploading”—is, unsurprisingly, a much more recent idea. The idea of digital survival after bodily

death is only about as old as electronic computer technology itself, with the earliest science fiction stories of mind uploading dating from the mid-1950s.<sup>1</sup>

As is often the case with ideas originating in science fiction, there are those who seek to transform the fiction into a reality. In the case of mind uploading (and related ideas such as the technological singularity, transhumanism, and posthumanism), among the most well-known proponents is author, inventor, and futurist Ray Kurzweil, currently a Director of Engineering at Google (a company with intense interest in artificial intelligence [AI] and related technologies). In several of his books (Kurzweil 1999, 2005, 2012) and elsewhere, Kurzweil has advocated work toward making immortality via mind uploading a reality in the present century. (And he himself is on an intense daily regimen of vitamins and other supplements in the hopes of surviving until the singularity.) Many with similar aspirations include self-styled “transhumanists” who seek technological transformation of human bodies and minds toward an eventual posthuman mode of being.<sup>2</sup> Relevant large-scale projects currently under way include the Human Connectome Project ([humanconnectomeproject.org](http://humanconnectomeproject.org)) funded by the U.S. National Institute of Health (NIH), and the Human Brain Project ([humanbrainproject.eu](http://humanbrainproject.eu)) and the Blue Brain Project ([bluebrain.epfl.ch](http://bluebrain.epfl.ch)), both funded by Switzerland’s École Polytechnique Fédérale de Lausanne. Of course, not all participants in these projects are upload advocates, but real-life upload advocates are enthusiasts for such projects.

As anyone at all familiar with metaphysics would expect, there’s room for metaphysical disagreement over whether mind uploading is even possible. Recent metaphysical work expressing optimism about the possibility of mind uploading includes Chalmers (2010) and Walker (2014). For samples of more pessimistic appraisals of the metaphysical possibility of mind uploading, see Pigliucci (2014) and Corabi and Schneider (2014). Although there are many components to the idea of mind uploading upon which pessimists and optimists may base their disagreements, there are two components that play central roles in the current debate. The first is the very idea of artificial intelligence, the idea that there could be any kind of digital computer—however complex or cleverly programmed, perhaps running a whole-brain emulation, perhaps not—that could support a genuine conscious mind. The other component of mind uploading over which metaphysical pessimists and optimists disagree is the proposition that some digital entities could be you. Even if one were to take an optimistic metaphysical view on the AI question, one might nonetheless remain a pessimist about whether an artificially conscious machine could be you, as opposed to a mere copy of you.<sup>3</sup>

1. The earliest descriptions of mind uploading in science fiction include Pohl (1955) and Clarke (1956). More recent appearances include the Dixie Flatline character from Gibson (1984). To my mind, the most philosophically interesting and thorough science-fictional treatments of mind uploading are in Egan (1994, 1997).

2. See various contributions in More and Vita-More (2013) for an overview. See also Koene (2014) and Sandberg and Bostrom (2008).

3. Other issues over which the pessimists and optimists might find themselves waging their dispute include, but are not limited to: (1) whether a person is a mind; and (2) whether minds in biologically instantiated human persons supervene on just their brains or also on extra-cranial bodily or even environmental items.

The debate about mind uploading under present consideration is a first-order metaphysical debate, a debate about which one might adopt various higher order—that is, meta-metaphysical—views. To taxonomize such views very coarsely, meta-pessimists about mind-uploading are disinclined to think that any winner can emerge in the first-order debate, whereas meta-optimists think that progress can be made toward winning the dispute. Meta-pessimists include skeptics who think that there are unknowable substantive facts under dispute. They also include those who hold the debates to be merely verbal or in some other sense empty. Meta-pessimists might be meta-pessimistic just about this metaphysical debate in particular, or metaphysical debates more generally. In contrast are the meta-optimists, who I presume to include the partisans in the debate. Why else would they engage in an exchange of reasoned arguments if they think there is no epistemic success that could be thereby achieved?

I myself feel a strong pull toward meta-pessimism about the mind-uploading debate. The pessimists and the optimists in the first-order metaphysical debate over mind-uploading seem to me to be locked in a stalemate, at least given the terms the debate is currently conducted in. Both sides make cases that, to my mind at least, are equally compelling.

The most compelling consideration favoring first-order pessimism about mind uploading is the very strong appearance that mind uploading would involve the creation of mere copies of a mind, as opposed to the survival of the original mind in a new, digital form. This appearance is strongest when we consider versions of mind uploading wherein the scanned brain isn't destroyed and continues to exist at the same time that the resultant digital mind exists. A similarly strong appearance of mere copying arises when we consider that the scan can be used as input to multiple digital emulations of the original brain. These separate emulations might be used to steer separate robot bodies inhabiting separate spatial locations. Since the original and the multiple copies can have different spatial locations and can have different futures, there's a strong appearance that instead of mind "uploading" allowing one person to persist across multiple locations and timelines, there are multiple persons thereby created by the procedure. On this pessimistic view, "mind cloning" might be a more apt name than "mind uploading" for the process under consideration.

The most compelling line of thought for first-order optimism about mind-uploading calls attention to (1) the models we already have for a metaphysics of survival through uploading and (2) the similarities that humans bear to those model cases. Regarding (1), consider the persistence conditions of, for example, William Gibson's novel, *Neuromancer*. The novel considered as the story itself is a thing to which we can ascribe properties (e.g., it's dystopian, it's cyberpunk), and its properties are distinct from those of the bound collections of printed pages I may purchase and set upon my shelf (e.g., some are hardcover, some are paperback). Though there may have been thousands of copies of *Neuromancer* printed, *Neuromancer* is just one novel, not thousands—it's the only novel that William Gibson wrote prior to his novel *Count Zero*. Those things of which there are thousands of copies bearing the title *Neuromancer* are not copies of the story itself. In order to copy the *story*, other requirements must be met, and another

print run won't alone suffice. One must instead, for example, plagiarize Gibson to produce a copy of the story in this latter sense of "copy." One might be tempted here to say that in talking about the story itself, we are talking about something abstract, or universal or repeatable, or something that exists at the level of types instead of the level of tokens. But terms such as "abstract," "universal," and the like are very much up for grabs and it's unclear whether introducing them creates more problems than they solve. Nonetheless, entities such as stories, poems, and melodies are abstract in that they have persistence conditions that are different from, say, some particular collection of particular particles arranged into paper and ink.

We might similarly come to regard a human person as abstract. Much of what we think, want, and experience is abstract. I can think that there's a dog chasing a cat without there being some particular dog or particular cat that I am thereby thinking about. As Quine (1956) points out, the desire I express in saying "I want a sloop" can just be me wanting relief from slooplessness without there being some certain sloop that I want. Regarding experiences and "what it is like" to have them: I can experience a patch of red on separate occasions, and what it is like to have the experience on the one occasion may be exactly like what it is like on the other occasion. Tye (1995) characterizes all phenomenal character as "abstract" in this sense. If what matters for having my mind is something that can be characterized as abstract in these ways, the possibility opens of a deep analogy between a human life and the story of a novel. The possibility opens that the relation between a single person and her multiple brain emulations or digital backups is importantly similar to the relation between the single story which is Gibson's *Neuromancer* and the thousands of copies that have been printed and sold since 1984. Metaphysically speaking, then, survival through mind uploading is no more an impossibility than is Gibson emailing his latest novel to his publisher.<sup>4</sup>

While I find this line of thought in favor of mind uploading compelling, I find it no more compelling than the cases that have so far been made against uploading. Nor do I find it any less compelling. The debate seems to me to be stuck at an impasse. However, it is not my aim here to establish that it is indeed an impasse. Instead I want to offer what I think is an interesting possible end run around the first-order metaphysical debate.

I draw my inspiration here from a similar impasse and its "resolution" that befell characters in Egan (1997). In Egan's tale, the future descendants of present humans divide into three groups: First are the Citizens, mind-uploads and their software descendants inhabiting virtual worlds in extra-planetary mainframes known as the polises. Second are the Gleisners, software minds inhabiting robot bodies on the moon and other airless off-Earth locations, who value real-world

4. Against the suggestion that people are abstract and non-concrete in a way that would make uploading survivable, Corabi and Schneider (2014) offer as their sole objection that it "violates deep intuitions we have about the nature of persons, and hence a view leading to these implications should be discarded" (pp. 139–40). But absent an independent means of verifying the reliability of metaphysical intuition, I see no reason to put much stock in such a consideration.

interaction much more than do the Citizens. Third are the Fleshers, biological Earth-bound descendants of present humans who allow alterations to their bodies via genetic engineering, but refuse to partake in the cybernetic extremes embraced by the Citizens and the Gleisners.

Crucial here is a conflict between the Citizens and the Fleshers, a metaphysical conflict that comes to a head when it is discovered that a wave of destructive radiation is headed toward the Earth from a distant cataclysm involving a pair of colliding neutron stars. The Citizens and the Gleisners have a chance at escaping the impending destruction of all life on Earth, but there's no hope for any biological life in the Solar System to either withstand the wave or make a timely escape, at least, not without abandoning their fleshy forms. Citizen envoys contact the Fleshers, offering them one last chance at mind uploading. The Fleshers refuse, on the grounds that what's being offered isn't a form of survival anyway. The Fleshers stay on Earth and meet their fate, as the wave of radiation destroys all of Earth's biological inhabitants. There is a sense here in which the Citizens (and the Gleisners) win the argument against their metaphysical opponents. But they do not win by convincing the Fleshers. Instead, the argument is won when the advocates of the contrary view go extinct.

Below I develop an argument that believing in the survivability of a mind uploading procedure conveys value to its believers that is assessable independently of assessing the truth of the belief. Regardless of whether the first-order metaphysical belief is true, believing it conveys a kind of Darwinian fitness to the believer. Of course, a further question remains of whether having that Darwinian property can be a basis—in a rational sense of being a basis—for one's holding the belief. I'll also make some remarks in the present article toward answering that latter question.

The arguments that follow depend on certain stipulations. Let us stipulate the following as a possible technology: A human brain can be scanned with such a high level of resolution that its current global state down to the molecular level can be recorded digitally, and further, the digital record can serve as an input to a whole-brain emulation that interacts with a highly realistic robotic body or high-resolution virtual body. Let us stipulate further that the scanning process is destructive in that the original brain is destroyed in the process. One final stipulation: The resultant entity would pass a Turing (or Turing-ish) test for being you—its verbal and other behaviors are indistinguishable from your own.

Of course, these stipulations make all sorts of assumptions about the *empirical* feasibility of mind-uploading technology. I do not mean to make light of such issues, but I do want to focus here on the distinctively metaphysical aspects of the debate. What's important, then, is that the stipulations do not beg any of the questions over which there is metaphysical dispute. Note that these stipulations leave open whether the entity that results from a mind-upload procedure is the survival continuant of the biological entity that was scanned. First-order optimists say (or bet) that it is. First-order pessimists say (or bet) that it is not. I turn now to sketch what I mean by calling the optimists more “metaphysically daring” than the pessimists.

## 2. HUMAN SURVIVAL STRATEGIES, POSTHUMAN SURVIVAL STRATEGIES, AND A SPECTRUM OF METAPHYSICAL DARING

Here's a human survival strategy: Keep your body healthy and alive and do not allow the destruction of your brain. This strategy has much to recommend it, and has worked for many humans still alive today. This is a strategy for the survival of individual humans. But another interest one might label "human survival" is interest in the survival of humanity as a whole. Surviving as a species for the indefinite long run will involve confronting contingencies such as a massive asteroid impact or the eventual death of the sun as it depletes its hydrogen reserves. Becoming a species able to cope with such "existential risks" (Bostrom 2013) may mean adapting our individual physical forms to better weather climate changes or to harden and shrink our bodies to make them better suited to space colonization. If the changes are sufficiently extreme, what survives may be less human than posthuman. Thus do some survival strategies become posthuman survival strategies. Along the way, modifications not only in outward physical form may be adaptive; so may modifications in metaphysical attitude.

Before unpacking what I will be calling the attitude of "metaphysical daring," let us first reflect a bit on the general notion of daring, especially on how daring relates to degrees of certainty about possible beneficial and costly outcomes. Consider some possible course of action one might contemplate undertaking, like for instance parachuting from an airplane or investing in some company. There are three crucial parts to what being daring involves. When one is daring or acts daringly, (1) one risks harms that are otherwise avoidable in the pursuit of (2) gains that would otherwise not be achievable, and further, (3) one does so without knowledge of whether the harms or the benefits will actually obtain. Regarding (1): If there were no foreseeable risk of harm in, say, parachuting from an airplane, or investing thousands of dollars in some startup, then the action in question would be no more daring than is brushing one's teeth or turning the page in a book. Regarding (2): if one risked harm, but there was no foreseeable gain in doing so—not either, for example, a financial gain, or a gain in exciting sensations, or fun memories, or cool stories to tell one's friends—then the action in question seems in no way rationally defensible. And if the action in question is irrational, we wouldn't laud it as daring but would instead dismiss it as foolhardy or worse. Regarding (3): The more the actor is subjectively certain that (a) the harm risked will occur or (b) that the gain sought will occur, then the actor is not so much daring and as instead either (a) merely self-sacrificing or (b) merely prudent.

There is a fourth crucial component to the general notion of daring, and that is the fact that there are degrees of daring—one person can be more daring than another person, and a given particular action can be more daring than another action. What determines variations of degrees of daring are variations within one or more of the first three components of the account so far sketched (with variations in the first factor—willingness to risk harm—likely being the most determinative of variations in daring). If Juan's jumping out of an airplane is more daring than Sarah's investing in pork futures, then there is some degree of difference between them with respect to one or more of the following factors: (1) their

respective estimates of amount and likelihood of harm; (2) their respective estimates of amount and likelihood of gain; and (3) their degrees of certainty about the obtaining of portions of the first two factors.

There is no doubt much more complexity that would need to be spelled out to give a full account of daring, but this coarse sketch should suffice here.<sup>5</sup> I'm not here interested in a full conceptual analysis of daring, but instead in inventing a useful notion of *metaphysical daring*. Toward that end, the main things I want to draw from this discussion of the general notion of daring are its admitting of degrees and its relations to the ideas of cost, reward, and ignorance. I turn now to the specific notion of metaphysical daring.

The scenario of having one's brain destructively scanned to be replaced by a digital whole-brain emulation is a metaphysically terrifying scenario. It's terrifying to the degree that it might involve your demise. It's metaphysical in that metaphysical views are in play about whether it would constitute your demise, or instead constitute a way of forestalling it by allowing survival through backup.

There are other hypothetical scenarios in this neck of the woods that we can rank with respect to the degrees of metaphysical terror they might inspire, and correlative degrees of daring it would take to overcome the terror. Consider, for example, teleportation, an oft discussed hypothetical scenario in the philosophical literature on the metaphysics of personal identity (Nozick 1981; Parfit 1984). One way that people imagine this technology as working is as similar in its basic principles to a fax machine with a built-in paper shredder. Suppose Jones wishes to travel from Earth to Mars without spending the money and time it would take to transport the mass of his body at sub-light speeds across the solar system. He contemplates utilizing the new teletransportation service he saw recently advertised. But he has certain misgivings. As is made clear in the advertisements, the technology at the sender station on Earth will destructively scan Jones's body. The scan will record information about the number and arrangement of molecules in Jones's body, but those very molecules will not be transported to Mars. Jones's body will be "dematerialized" on Earth—his molecules ripped apart and stored in an Earth-bound tank for later use. What instead travels to Mars are radio waves containing information gained from the scan of Jones. When the radio transmission arrives at the receiver station on Mars, there's a store of molecules waiting there on Mars and arranged in conformity with the information specified in the received transmission.

So as not to beg any of the relevant metaphysical questions, call the assembled entity at the Mars receiver station Schmones. The metaphysical debates about Schmones concern whether Schmones has a mind and whether Schmones is Jones. Alternately, we can state the question in terms of survival: Does Jones survive and now live on Mars? In other words: Is Schmones a survival continuant of Jones?

5. One thing that might be added to further elaborate the basic account of daring is reference to a social context such that even a person with high confidence in success is daring if the embedding social context is largely comprised by members with a low confidence. I'm grateful to Eric Schwitzgebel (personal conversation) for the suggestion.

The following points are agreed upon by the disputants of the metaphysical debates about mind and identity in connection with teletransportation: There's no task that Jones is able to do that Schmones cannot do and vice versa, and this includes tasks concerning verbal behavior. Any set of questions that Jones would answer correctly are questions to which Schmones would respond with the same utterance types. (The utterances are here typed independently of meaning—we are here bracketing whether Schmones's utterances have the same *meanings* as Jones's utterances.)

For some, this hypothetical teleportation is an intolerably terrifying proposal. From the point of view of the metaphysically timid, this is not so much a form of transportation as it is a form of suicide—Jones life ends in the so-called sender station and Schmones is thus not Jones's continuant. In contrast to the metaphysically timid are the metaphysically daring. They embrace teletransportation as an enhancement of their power. Despite not knowing for certain whether the thing appearing in the receiver station is their future continuant, they boldly hold that it is, or at least they boldly bet on it. The benefit they seek is an enhancement of the speed and range of their travels. A further potential benefit of the basic technology of teleportation is to use it as a way of creating backups of an entire person. If, after arriving at the receiver station, our traveler meets a fatal calamity, as long as the relevant information is still in the memory buffer at the station, a new one can be "printed out."

An even more daring metaphysical position than the one described in connection with teletransportation is one that embraces whole-body emulation as a digital means of survival. Suppose we take the information gained from the scan of Jones, and instead of constructing the molecular doppelgänger Schmones, we utilize the information as input to a computer simulation of Jones's predestruction body. Call this software entity Digital Schmones. It is even more daring to hold that Digital Schmones is a continuant of Jones than it is to hold Schmones to be a continuant of Jones. Holding Schmones to be the survival continuant of Jones is a safer bet than holding Digital Schmones to be Jones's survival continuant. This is because Schmones is more similar to Jones than is Digital Schmones. In the case of Jones and Schmones, they have intrinsic physical constitution in common—they are constituted by the same number and types of particles, in the same relative arrangements. But in the case of Jones and Digital Schmones, since they have less in common, it is a greater leap—and thus a greater feat of metaphysical daring—to bet on the survival-continuity of Jones and his whole-body emulation counterpart.

We can generalize the above remarks as follows: The more a being differs from Jones, the more metaphysical daring it takes to regard that being as Jones's continuant. Just as it is more daring to regard Digital Schmones as a Jones-continuant than it is to so regard Schmones, it is more daring still to so regard an emulation of only Jones's brain. And it takes even more daring still to regard as a Jones-continuant a coarse-grained brain emulation that abstracts away from the molecular level and doesn't emulate any process at a grain finer than that of whole neurons. More daring still is to forsake scanning the brain of Jones, and to rely instead on what Steinhart (2014) calls *digital capture* and Chalmers (2010) calls *reconstructive uploading*, wherein data about Jones are gleaned from his Twitter



timeline, Facebook status updates, and so forth, and used as input to a computer program that extrapolates to digitally reconstruct his mind.

In each of the series of hypothetical cases I have been describing, the resultant entities are increasingly more different from the original entities, where the original entities are flesh-and-blood human beings. In each of the hypothetical scenarios, with the greater departure from the properties of the original comes a greater risk that something essential to survival-continuity has been left out. However, it makes little sense to risk big without some chance of winning big. Let us turn, then, to this question: What are the potential gains awaiting the highly metaphysical daring?

### 3. DIGITAL DARWINISM

There are several advantage-bearing traits that go along with being metaphysically daring, traits that can be inherited by upload “progeny,” who are likewise daring. However, some care will be needed here in describing these inheritable traits. I want to describe the upload progeny as being capable of inheriting certain traits without thereby begging any metaphysical question that defines the first-order metaphysical dispute over uploading. It won’t do to describe the upload as sharing the metaphysical *beliefs* of their daring biological parents, since some parties to the metaphysical dispute aren’t willing to grant that the so-called uploads have all of the same beliefs or perhaps any beliefs at all. It will be useful, then, to define a notion of *quasi-belief*.<sup>6</sup>

Consider, once again, biologically brained and embodied Jones and his upload counterpart Digital Schmones. Jones and Digital Schmones have exactly similar causal dispositions at some coarse level of grain. Jones’s belief that snow is white plays a causal role in Jones and is causally related to certain of Jones’s “snow is white” utterances. We can identify in Jones a node in a coarse-grained causal network that corresponds to, but may or may not be identical to, Jones’s belief. Corresponding to Jones’s belief that snow is white is Digital Schmones’s quasi-belief that snow is white. This quasi-belief of Digital Schmones’s is an element in Digital Schmones’s coarse-grained casual network which plays the same role in relation to Digital Schmones’ “snow is white” utterances and so on, as Jones’s belief plays relative to his own such utterances and so on. Now, if certain views in the metaphysics of mind are correct, namely, certain versions of functionalism (especially certain internalist or individualistic versions of functionalism), then quasi-belief and belief are one and the same. However, the definition of quasi-belief is here explicitly stipulated in such a way as to be neutral with respect to the metaphysical controversies in this neck of the woods.

While it is a point of controversy whether any of Digital Schmones’s quasi-beliefs are also beliefs, it is entirely uncontroversial that each of Jones’s beliefs is also a quasi-belief. Despite there being fine-grained levels of causal detail that Jones has and Digital Schmones lacks, there is a level of grain at which they are the

6. While I take my inspiration here from the notion of quasi-memory due to Shoemaker (1970), I am not taking on board all that his notion involves.

same, and it is in virtue of that similarity they have the same quasi-beliefs. Of particular interest to the present discussion are the beliefs of Jones's that constitute his overall level of metaphysical daring. Suppose that Jones has a high level of metaphysical daring and believes, or believes that it is highly likely, that he will survive a destructive upload. It follows, given the definition of quasi-belief, that Jones also quasi-believes that proposition.

It follows from the way we are defining quasi-belief that Jones's upload-counterpart Digital Schmones will inherit all of Jones's quasi-beliefs. Of particular interest is the consequence that Digital Schmones will inherit quasi-beliefs that will make Digital Schmones also daring. Perhaps we should instead say that Digital Schmones is *quasi-daring*. Alternately, we can alter the definition of metaphysical daring to be explicitly defined in terms of quasi-belief so that, for example, what it means for Mary to be more metaphysically daring than Sue is that Mary and Sue differ in certain propositions that they quasi-believe. For example, Sue quasi-believes there's more of a chance of surviving destructive mind uploading than does Mary.

Now we are in a position to compare the various degrees of survival advantage that go along with different quasi-beliefs, especially those quasi-beliefs that are constitutive of differing degrees of metaphysical daring (or metaphysical quasi-daring). Consider, for example, the quasi-belief that one can extend one's chances of survival by allowing digital backups of oneself to be made. Compare this to the quasi-belief that one will not survive such a procedure. Consider two entities, Mary and Bob. Mary quasi-believes that she can maximize her chance of survival by creating digital backups of herself, and proceeds to create such software entities. We can expect her backups to share that quasi-belief, and at some future date, when Mary is "restored" from a backup, the restored "Mary" will continue, resources permitting, to engage in backing-up behaviors. Bob, on the other hand, does not quasi-believe in the possibility of survival through backup. In contrast, he has the quasi-belief that one cannot survive via such a procedure and thus does not consent to have any backups made.

The applicability of evolutionary theory is quite obvious here. Mary's quasi-belief leads to a behavior that results in many more copies of that very quasi-belief and that quasi-believer. Mary's various backups, as well as the backups of her revival counterparts, constitute a kind of superior reproductive fecundity—Mary's backups and as well as her backups' backups all can be counted as Mary's digital progeny. Mary's metaphysically daring quasi-beliefs make her and her progeny that share them more fit than Bob and his ilk. Bob and other metaphysically timid individuals have quasi-beliefs that rate low on the scale of metaphysical daring, and as such, are less likely to avoid the existential risks that threaten the wholesale extinction of such metaphysical views.

In backing up multiple copies in multiple locations, Mary increases the chances of the persistence of her quasi-beliefs despite whatever damaging events might strike any particular location. Bob, being much more timid and not consenting to back up, is thus limited to a single place at a time and may not survive any calamities that strike there. Further, Mary's backups might be distributed across different kinds of media, some invulnerable to some kinds of destructive event,

while others have different invulnerabilities. This increases the survivability of Mary's quasi-belief and her quasi-believer "offspring" still further.

Depending on how daring upload-optimistic Mary is, further dimensions of advantage may open to her and/or her upload counterparts. If Mary views the requirements on her digital survival as requiring the emulation of her whole body or brain down to the molecular level, she is going to choose a data plan for backup services that is much more expensive than if she regards survival as requiring only information about her brain at a much coarser grain of resolution. If, on the other hand, she holds her survival requirements to require much smaller portions of her body than her whole nervous system, or much coarser levels of grain than the molecular or sub-cellular level, then the memory burden of each backup is smaller, and she can store more backups.

Different degrees of metaphysical daring can be described as different bets about how much must be retained to maintain personal survival. Some, if not all, of these bets can be stated in terms of amounts of information that must be preserved by survival-continuants. Moravec (1998) offers  $10^{45}$  bits as an estimate of the information contained in a whole human body considered at the lowest levels of grain allowable by contemporary physics (those characterized in terms of the "Bekenstein bound"). Focusing on just the human brain, and at a coarser level of grain, Moravec offers the much lower estimate of  $10^{15}$  bits. Rothblatt (2013) gives a lower estimate still, describing a coarse-grained record that constitutes "a digital file of a person's mannerisms, personality, recollections, feelings, beliefs, attitudes and values." On this latter view, all that's required is a single terabyte ( $= 8 \times 10^{12}$  bits).

Memory is one of the main precious resources in a digital ecology. Fitter digital creatures are likely to be those that take up less memory. In artificial life simulations such as those by Ray (1992) and Sayama (1999, 2004), evolutionary competitions between reproducing and mutation-prone digital organisms give rise to the spontaneous emergence of viable descendant populations that were digitally smaller than their forebears. The smaller organisms were poised to have more copies in an allotted space than their larger competitors. In Ray's famous *Tierra* simulation, one of the ways that this digital smallness was achieved was via a form of parasitism. The parasites could exploit the reproductive mechanisms of their host organisms, and thus could themselves be smaller in not needing such mechanisms of their own. In Sayama's *Evoloop* simulations, the mutations that led to the more dominant smaller species did not give rise to parasitism, but nonetheless gave rise to self-reproducing organisms that were able to fill a given space with a greater number of copies than their larger competitors. Metaphysically daring posthumans competing in digital ecologies may find similar advantages in betting on survival despite reductions in digital size.

Consider a comparison between two posthuman software entities, Mary and Jane. They are both more metaphysically daring than Bob, a biological entity who refuses to submit to a destructive uploading procedure. However, Mary and Jane, let us suppose, still differ in degrees of daring. They differ in their opinions or quasi-opinions of how fined-grained the resemblance has to be between two entities in order for one to count as the survival continuant of the other. Mary is more

metaphysically daring than Jane—Jane holds that a simulation of a brain has to simulate it down to the molecular level; Mary, in contrast, holds that a lower resolution of similarity, a coarser grain than the molecular grain, suffices for survival via digital backup. Given some finite quantity of computer memory, the more daring Mary will be able to store more backups than the less daring Jane. More backups means a better chance of survival or quasi-survival, especially if the backups are stored in multiple places and in multiple media. So, Mary's greater degree of metaphysical daring brings with it a greater a degree of fitness that will be inherited by her upload progeny.<sup>7</sup>

#### 4. A QUESTION OF VALUE

Assuming that what I have said so far is true, there may nonetheless remain an open question about what's desirable or rational to value for present-day biologically constituted humans. I assume that you, dear reader, are concerned with your own survival, and I also assume that you might be reasonably uncertain whether to undergo a mind-uploading procedure. Here my hunch is that the best way to respond to your doubts is to draw attention to possible analogies to values you might already hold, and further, are not unreasonable to hold:

- Good outcomes for your biological offspring, and their offspring, including future generations beyond your own lifetime.
- Good outcomes for your adopted children, and their own children, even if they themselves adopt.
- Good outcomes for your own intellectual products even beyond one's own life.
- Good outcomes for your species considered as a whole.
- Good outcomes for living things considered generally, thus including species not your own.
- Good outcomes for a team with which you identify or for which you root, and do so on the basis of the team's past track record and future likelihood of being the winning team.

Analogies between the above-listed values and the value stance of a metaphysically daring upload-optimist allow such a value stance to be intelligible, and thus, intelligibly rational. Present-day upload-optimists may themselves see their advocacy for uploading as grounded in values importantly analogous to goals in the above list. They may, for instance, see an upload-dominated technological posthumanity as a desirable direction toward which to steer the future of human

7. I'm grateful to Eric Schwitzgebel (personal communication) for suggesting the following worry: My argument here assumes the sort of notion of fitness widespread in artificial life research whereby it is sufficiently abstract to apply to digital entities as well as ordinary biological life. Against an imagined objector who insists that fitness only applies to the naturally biological, and that anyone who uploads prior to biologically reproducing is thus unfit, I would respond that the argument can just as well go through by shifting from a notion of fitness to a notion of quasi-fitness. This is not, however, to grant that the objector is right about what the realizers of the property of fitness are.

evolution. They may, as a self-identified transhumanist or posthumanist, view their own efforts toward actualizing mind-uploading technology as contributing to bringing about the existence and thriving of the ultimate winning team.

The upload optimists in Egan (1997)—the virtual Citizens inhabiting the computer polises, and the Gleisners whose software minds inhabit robot bodies—win the evolutionary race to become the ultimate descendants of humanity, and do so because of their metaphysical daring. Their Flesher competitors go extinct as a result of their metaphysical timidity. Of course, to assert that the survival of the Citizens and Gleisners *constitutes* the survival of humanity is to beg a metaphysical question against the Fleshers and other upload pessimists who would hold that Flesher extinction *just is* the extinction of humanity. Nonetheless, not only are the victors often in a position to rewrite history; they may also find themselves in a position to rewrite metaphysics.<sup>8</sup>

## REFERENCES

- Bostrom, Nick. 2013. "Existential Risk Prevention as Global Priority." *Global Policy* 4(1): 15–31.
- Chalmers, David J. 2010. "The Singularity: A Philosophical Analysis." *Journal of Consciousness Studies* 17: 7–65.
- Clarke, Arthur C. 1956. *The City and the Stars*. London: Frederick Muller.
- Corabi, Joseph, and Schneider, Susan. 2014. "If You Upload, Will You Survive?" In *Intelligence Unbound: The Future of Uploaded and Machine Minds*, ed. Russell Blackford and Damien Broderick, 131–45. Oxford: Wiley Blackwell.
- Egan, Greg. 1994. *Permutation City*. London: Millenium.
- . 1997. *Diaspora*. London: Millenium.
- Gibson, William. 1984. *Neuromancer*. New York: Ace Books.
- Koene, Randal. 2014. "Feasible Mind Uploading." In *Intelligence Unbound: The Future of Uploaded and Machine Minds*, ed. Russell Blackford and Damien Broderick, 90–101. Oxford: Wiley Blackwell.
- Kurzweil, Ray. 1999. *The Age of Spiritual Machines*. New York: Viking.
- . 2005. *The Singularity is Near*. New York: Viking.
- . 2012. *How to Create a Mind: The Secret of Human Thought Revealed*. New York: Viking.
- Moravec, Hans. 1998. "The Senses Have No Future." In *The Virtual Dimension: Architecture, Representation, and Crash Culture*, ed. John Beckmann, 84–95. New York: Princeton Architectural Press.
- More, Max, and Vita-More, Natasha, eds. 2013. *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*. Oxford: Wiley Blackwell.
- Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Pigliucci, Massimo. 2014. "Mind Uploading: A Philosophical Counter-Analysis." In *Intelligence Unbound: The Future of Uploaded and Machine Minds*, ed. Russell Blackford and Damien Broderick, 119–30. Oxford: Wiley Blackwell.
- Pohl, Frederik. 1955. "The Tunnel Under the World." *Galaxy Science Fiction* 9(4): 6–37.

8. For discussions of early versions of this work, I thank Eric Schwitzgebel, Howie Wettstein, Ken Williford, Eric Steinhart, Richard Brown, Susan Schneider, Scott Bakker, and Malcolm Nicholson. I also thank the audiences at presentations of earlier versions to the departments of English and of Philosophy at the University of Texas, Arlington, and at "A Night of Philosophy" at the French Embassy in New York City.

- Quine, Willard V. O. 1956. "Quantifiers and Propositional Attitudes." *Journal of Philosophy* 53: 177–87.
- Ray, Thomas S. 1992. "An Approach to the Synthesis of Life." In *Artificial Life II*, ed. Christopher G. Langton, Charles Taylor, J. Doyne Farmer, and Steen Rasmussen, 371–408. Redwood City, CA: Addison-Wesley.
- Rothblatt, Martine. 2013. "We Are the World: Inviting Everyone Onboard the 100YSS is Practical and Will Help to Ensure Its Success." Retrieved April 13, 2013, from <<http://www.kurzweilai.net/we-are-the-world-inviting-everyone-onboard-the-100yss-is-practical-and-will-help-to-ensure-its-success>>.
- Sandberg, Anders, and Bostrom, Nick. 2008. "Whole Brain Emulation: A Roadmap." Technical Report #2008-3. Future of Humanity Institute, Oxford University.
- Sayama, Hiroki. 1999. "A New Structurally Dissolvable Self-reproducing Loop Evolving in a Simple Cellular Automata Space." *Artificial Life* 5(4): 343–65.
- . 2004. "Self-protection and Diversity in Self-replicating Cellular Automata." *Artificial Life* 10(1): 83–98.
- Shoemaker, Sydney. 1970. "Persons and Their Pasts." *American Philosophical Quarterly* 7(4): 269–85.
- Steinhart, Eric. 2014. *Your Digital Afterlives: Computational Theories of Life after Death*. New York: Palgrave Macmillan.
- Tye, Michael. 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
- Walker, Mark. 2014. "Uploading and Personal Identity." In *Intelligence Unbound: The Future of Uploaded and Machine Minds*, ed. Russell Blackford and Damien Broderick, 161–77. Oxford: Wiley Blackwell.