

International Journal of Machine Consciousness
 Vol. 4, No. 1 (2012) 1250014 (15 pages)
 © World Scientific Publishing Company
 DOI: 10.1142/S179384301250014X



WHY UPLOADING WILL NOT WORK, OR, THE GHOSTS HAUNTING TRANSHUMANISM

PATRICK D. HOPKINS

*Department of Philosophy,
 Millsaps College, 1701 North State Street,
 Jackson, MS 39210, USA
 hopkipd@millsaps.edu*

Transhumanists tend to have a commitment to materialism and naturalism but nonetheless pursue goals traditionally associated with religious ideologies, such as the quest for immortality. Some hope to achieve immortality through the application of a technology whereby the brain is scanned and the person “uploaded” to a computer. This process is typically described as “transferring” one’s mind to a computer. I argue that, while the technology may be feasible, uploading will not succeed because it in fact does not “transfer” a mind at all and will not preserve personal identity. Transhumanist hopes for such transfer ironically rely on treating the mind dualistically — and inconsistently with materialism — as the functional equivalent of a soul, as is evidenced by a carefully examination of the language used to describe and defend uploading. In this sense, transhumanist thought unwittingly contains remnants of dualistic and religious concepts.

Keywords: Copying; immortality; Hans Moravec; materialism; personal identity; soul; transhumanism; uploading; whole brain emulation.

1. Introduction

One of the fascinating and appealing things about transhumanism is the way it tends to embrace materialism, naturalism, and science but at the same time espouses goals that have traditionally been associated with religion. Talk of transcendence and enlightenment [Alexander, 2003], spiritual machines [Kurzweil, 1999], omega points and singularities [Kurzweil, 2005], universal consciousnesses [Moravec, 1999], moral perfection, transubstantiation [Alexander, 2000, 2003], and escape from a merely earthly existence abound in discussions of transhumanism and the technologies transhumanists find so promising — and there is no inconsistency here. The kinds of states, beings, qualities, experiences, and goals entertained in these discussions are all supposed to rely on perfectly ordinary materialist and naturalist principles. It is the transformation of matter and energy that is expected to produce the changes in cognition, embodiment, and society that will carry us forward to these religion-analog futures. No magic or supernaturalism is involved.

P. D. Hopkins

Certainly not least among these (traditionally religious) goals is immortality — the age old desire to escape the clutches of death and decay, to live beyond the ordinary limits of human lifespan, to face a future of unlimited experience and exploration without the certainty that we eventually simply stop. The goal is not only immortality itself, but also an immortal existence that is free of the confines of the frail and limited human body. Living forever is desirable, but being trapped or imprisoned in the type of body we were born with would be depressingly limited. Living forever in a form that allows far greater freedom and ability (to achieve knowledge, experience, transcendence) would be even better.

While various religious ideals may promise freedom and immortality in the form of freeing the soul from the body, or becoming one with God, or entering a perfect heaven, transhumanism avoids the questionable supernatural metaphysics of such approaches by offering purely materialist and naturalist methods for achieving immortality. Though such methods include technologies such as cryonics (<http://www.alcor.org/>) and engineered negligible senescence (<http://www.sens.org/>), the most permanent and far-reaching method is called “uploading.”

2. The Ghostly Language of Uploading

To understand what uploading is, let us examine a number of definitions and characterizations of the proposed process, and in doing so, pay close attention to the language used. First, I will begin with relatively short and simple descriptions:

- Uploading is the transfer of the brain’s mindpattern onto a different substrate (such as an advanced computer) which better facilitates said entity’s ends. Uploading is a central concept in our vision of technological ascension... [Kadmon, 2003]
- In transhumanism and science fiction, mind transfer... refers to the hypothetical transfer of a human mind either into a computer or other non-human receptacle, or from one human body to another. [Knowledgerush, 2009]
- Mind uploading is a radical form of human enhancement, whereby the human mind is transferred from the vulnerable organic medium of the brain to a computer system of some kind. [Human Enhancement and Biopolitics, 2009]
- In transhumanism and science fiction, mind transfer... refers to the hypothetical transfer of a human mind, body and environment to an artificial substrate. [Fact-Archive.com, 2005]
- Mind uploading, sometimes called whole brain emulation, refers to the hypothetical transfer of a human mind to a substrate different from a biological brain, such as a detailed computer simulation of an individual human brain. [Sentient Developments, 2009]
- Involving the transference of a mind from biological brain to computer hardware — or, for that matter, any other substrate... mind uploading is a tenet of transhumanist hopes and science fiction. [Keim, 2009]

Why Uploading Will Not Work

- The last stage, mind uploading, leads as logically on from brain emulation as brain emulation does from neuroprosthetics. Once it is possible to emulate a brain, it should be possible to transfer the mind within that brain from one set of hardware to another. That is, from a collection of neuroprosthetics, to a computer-based existence, taking the mind, and uploading it into the machine. [Virtual Worldlets Network, 2006]

As described, the basic idea in uploading is that technology will allow us to transfer a mind from a brain to a computer (or other substrate). Various ways are projected for how this might operate — using a microtome scanning procedure, using nanotechnology to record and reproduce the exact functional architecture of a brain, or using various other forms of scanning technology to record and replicate neural organization. The specifics of the methodology are not that important for the issue at hand. Let us grant that some or all of these technologies and methods will work, in the sense that they will allow us to scan the architecture of the brain down to the smallest level and reproduce that architecture elsewhere. The important question for our purposes here is not whether the technical apparatus will work, but rather, what will actually be accomplished even if the technology functions perfectly.

Now look carefully at the language in these descriptions. An important point researchers on metaphor have made is that metaphor is such a pervasive part of our ordinary language use, we often fail to see that a metaphor is being used at all. For example, as Lakoff and Johnson argue, when we talk about someone “winning an argument” or having “weak points in his position” or someone’s criticisms being “on target,” we often do not recognize that we are using war metaphors for argumentation [Lakoff and Johnson, 1980]. A criticism cannot literally be “on target” and a belief cannot literally be “weak” and an argument cannot literally have “holes” in it — these are metaphors. Similarly, when we talk about talking itself, we may wonder if we “got our idea across to her” or correctly “put our ideas into words” or we may suspect that someone’s speech “carried no meaning.” These are conduit and directional metaphors. An idea cannot literally be “put” “into” a word — words are not physical containers. We cannot literally get an idea “across” to someone — ideas are not objects that are moved around.

So when we examine the language used in descriptions of uploading, we should pay attention to whether metaphors are present, discerning what this reveals about the concept of uploading and what it is supposed to accomplish. Immediately noticeable is the central claim that the mind will be “transferred”. Again and again, the term “transfer” appears. Specifically, the idea is that by replicating the architecture of the brain in another substrate, we would be transferring the mind or the consciousness of that brain to the new physical system. But what does this language assume? First, the language uses spatial and motion metaphors. The mind is located “in” a particular place — some brain. Using technology, we will be able to move (“transfer”) the mind from “within” the brain it is currently located to another location — “onto” another substrate or “into” a computer or to another “receptacle”. Second, thinking of the

P. D. Hopkins

mind as “in” the brain suggests thinking of the mind in terms of substance — the mind is being treated as a thing, an object, something that is locatable and takes up specific space and that can therefore be moved from “inside” one thing to “inside” another.

Now of course, using metaphors is not necessarily bad. Problems arise, however, when we are not aware that we are using metaphors and when we are not aware of how metaphorical language can trick us into thinking incorrectly about concepts — particularly in treating something as literal that is only metaphorical. This is the problem going on with the concept of uploading. The mind is treated as a substance and an object that can be moved around different locations and deposited into different physical receptacles. But is this an accurate description of the mind? Do transhumanists believe that the mind is an object that is literally housed “inside” a brain and through technology can be “moved” from one “receptacle” to another? Not according to the materialist and naturalist worldviews transhumanists typically espouse. Materialists and naturalists may think of the mind as produced by brain activity or perhaps as some brain activity itself (these are complicated philosophy of mind issues), but they would not typically hold that the mind is a discrete object that sits in the container of the brain. In fact, if we look for positions that have held the view that minds or consciousnesses are actually substantial objects that have location and can be moved from one body to the next, we do not have far to look — only so far as popular religion. What else has been understood as a substantial consciousness “inside” a body that can be “moved” to another body or space intact? — a soul. A spirit. A ghost. Ironically, uploading enthusiasts by and large seem to be relying on a dualist theory of mind (which they normally explicitly reject) informed by religious metaphysics and beliefs in transmigratory souls and displaceable ghosts. The mind is being treated just like a soul.

At this point a reader may object that I am making too much of the language. Proponents of uploading do not really think the mind is a substance that can be moved. They probably think that the mind is the product of specific material architecture and activity. They are not really treating minds as ghosts even if the popular descriptions of uploading use all these spatial and motion metaphors.

The problem with this objection is that something utterly crucial depends on the metaphors of location and motion and the operational parallels to ghosts and souls. In the descriptions of uploading, the very core of the concept is that a specific mind is transferred from a brain to a computer. Such descriptions do not simply make the materialist point that minds are the result of physical activity, rather they make claims about the preservation of identity. They say — and this is whole point of uploading, the whole point of its connection to immortality and transcendence — that a specific, intact mind can be “transferred” (moved) from one embodiment to another. If one wants to say that, in spite of the language used, we should not take metaphors of transfer and movement too seriously, we are still left with the critical, literal, claim that the mind in the computer is the same mind as the mind in the

original brain. If that is not the case, then uploading is pointless in terms of immortality or enhancement or transcendence. If the specific mind that is “in” or produced by a brain is not the very same specific mind that is “in” or produced by the computer, then immortality has not been achieved, a person’s life has not been saved, and uploading fails to satisfy its original promise.

Now, the reason that metaphorically treating the mind as a locatable, moveable substance is attractive is precisely because it seems to solve (or at least gloss over) the identity preservation problem. If the mind “in” the computer has simply been “transferred” from a brain, then it has undergone no substantial change and thus is easily understood to be literally the same, or “identical” mind that was in the brain. It would be like repotting a plant — you could transfer a plant from one pot to another, but the plant is still literally the same plant. Spatial and motion terminology smooths over the problem of identity preservation. In fact, the very hope for immortality actually requires these spatial and motion properties of minds not to be metaphors at all, but to be literally true.

However, since minds, according to materialist and naturalist principles, are not literally movable substances (like natural material objects such as plants or supernatural immaterial objects such as souls), the preservation of identity problem is not solved, even though our language may make it difficult to see. If the spatial and motion metaphors are in fact just metaphors and minds are not movable substances, then hopes for uploaded immortality fall apart.

Essentially, the problem with uploading is an only marginally updated version of the old philosophical problem of personal identity — under what circumstances can we say that a person remains the very same person even while undergoing change? The real question in uploading is whether uploading procedures maintain the identity of the specific mind throughout the process. The spatial and motion language of transference assumes it does, but at the cost of treating minds like ghosts.

3. The Metaphysics of Copying

Of course, even though simple descriptions of uploading are clouded by spatial and motion metaphors and by treating minds as substantial souls, proponents of uploading do not believe minds are literally being carried from one place to another. The “transference” of the mind from here to there is taken literally, but is supposed to be accomplished by emulation, simulation, or replication — all terms that boil down to the concept of copying. The assumption is that by copying the mind you simultaneously transfer the mind, maintaining identity throughout. Copying = transference.

While simplistic descriptions of uploading often make this assumption without discussing the problems of identity preservation at all, there are explanations of uploading that do address the concerns about personal identity. So, even if simplistic understandings of uploading may be naively or inaccurately worded, do these more complex understandings of uploading succeed in showing that the process can achieve

P. D. Hopkins

what we really want it to? In order to answer that question we need to look at both the language that is used in such discussions and look at the specific arguments that try to show uploading-as-mind-copying does maintain identity.

One of the seminal works (and most popular) on what comes to be called uploading is Hans Moravec's discussion in the "Grandfather Clause" chapter of *Mind Children* [Moravec, 1988]. In a section of the chapter called "Transmigration" (another obvious religious allusion), Moravec asks the question "Is there a way to get our mind out of our brain?" His answer describes a potential process in which a robot brain surgeon microscopically scans the layers of your brain, constructs a 3-D chemical map, writes a program modeling the neural tissue's behavior, and then installs and activates the program in a computer. Checking the functionality of the program by allowing you to switch over to the simulation periodically, you are able to fine-tune the program so that it matches what your original neural tissue could do — in terms of movement, sensation and cognition. As more and more of the simulation is tested, the brain cells originally responsible for those activities are removed. Eventually your entire brain is destroyed and your body dies, but your consciousness is described as having shifted perspective to the computer.

This kind of process is sometimes referred to as "destructive uploading" since the brain is destroyed as the simulation takes over. Moravec also describes other potential methods including electronic scans and cybernetically enhanced split brain surgeries. However, the specifics of the techniques that are employed are not that important, as long as they all operate by some principle of simulation or copying. I will assume that the technology is feasible and would perform as described. The question here is what is accomplished in this procedure? This is what Moravec thinks is accomplished:

- Though you have not lost consciousness, or even your train of thought, your mind has been removed from the brain and transferred to a machine [110].
- Ultimately your brain would die and your mind would find itself entirely in the computer [112].
- You may choose to move your mind from one computer to another that is more technically advanced... [112].
- The program can also be copied to a future equivalent of magnetic tape. Then, if the machine you inhabit is fatally clobbered, the tape can be read into a blank computer... [112].
- As a computer program, your mind can travel over information channels...[114].

First of all, notice that the spatial and motion metaphors are fully in use here. Minds are treated as things that can be "moved", that can be "in" computer or brains, and that can "travel." Even though we are also told that what is going on in these processes is that the mind is being copied (to tapes, computers, etc.) the idea is that the copying manages to "move" the mind from one place to another. Copying seems to accomplish the same thing as moving.

Now, on the face of it, this is a strange assumption. In many cases we do not think that the act of copying something “moves” it. We instead think that copying makes another object that is just like the original in aspects relevant to our needs — but we do not think the copy literally is the original. For example, if I copy a page from a book, I have not “moved” the page from the book to the printout paper exiting the copier. The page is still there and the printout is a new object that looks just like the page, and note, it does not make the slightest difference if I quickly or slowly destroy the page in the book as I make the copy — the copy is still just a copy. It is a new object, whether the original is destroyed or not. Why wouldn’t the same be true if my neural architecture was copied? The copy would be just like me, but would not be the very same thing as me. My mind would not have been “moved”. My mind would have just been copied — a new mind functionally just like mine, but new nonetheless.

Now perhaps this ordinary use of “copy” is a bad analogy because we are talking about personal identity here rather than identity of physical objects like pages and printouts. Perhaps there is something about personhood or personality that operates according to different rules than physical objects. This is worth considering, but notice two things — first, it depends on treating personal identity as something other than physical (does this conflict with materialism and move us toward substance dualism?), and second, it claims that copying will do for some phenomena what it will not do for others. The second point is the more directly relevant one for our purposes here and seems to be what Moravec has in mind. He argues:

- The idea that a human mind can be transferred to a new body sometimes meets the following strong objection from people who do not dispute the theoretical possibility: “Regardless of how the copying is done, the end result will be a new person. If it is I who am being copied, the copy, though it may think of itself as me, is simply a self-deluded impostor. If the copying process destroys the original, then I have been killed. That the copy may then have a great time exploring the universe using my name and my skills is no comfort to my mortal remains.” This point of view, which I will call the body-identity position, makes life extension by duplication considerably less personally interesting. I believe the objection can be overcome by acceptance of an alternative position which I will call pattern-identity. Body identity assumes that a person is defined by the stuff of which a human body is made. Only by maintaining continuity of body stuff can we preserve an individual person. Pattern-identity, conversely, defines the essence of a person, say myself, as the pattern and the process going on in my head and body, not the machinery supporting that process. If the process is preserved, I am preserved. The rest is mere jelly [116–117].

The claim here is that what makes a person the specific person they are is a specific pattern. Copying the pattern “preserves” it, therefore the person is preserved through copying even though the body may be destroyed. Now, at first, this view seems an improvement over the naïve metaphors used in the rhetoric of “moving

P. D. Hopkins

minds". It is acknowledged that what is really a mind is a pattern of activity and since the pattern can be copied, the mind can be transferred. But this pattern-identity theory of preserving personal identity in turn depends on the claim that copying "preserves" the pattern. That is, even granting that cognitive activity patterns are our "essence" and what makes us, us, unless we can make good sense of the idea that copying actually does "preserve" the pattern, we will not have secured the claim that the copied mind is the very same mind as the original. Identity of minds cannot be preserved through copying unless identity of patterns can be preserved through copying. The word "preserve" is a strong and crucial term here. Unless the pattern of the copied mind is in fact the very same pattern (identical, not just exactly similar), no progress has been made. So how does Moravec defend the necessary claim that the pattern is literally preserved? He makes several points and analogies, which I will analyze in turn.

- Let me explore some of the consequences of the pattern-identity position ... A matter transmitter might scan an object and identify its atoms or molecules one at a time, perhaps removing them in the process. The identity of the atoms would be transmitted to a receiver, where a duplicate of the original object would be assembled in the same order from a local supply of atoms. . If solid objects, why not a person? Just stick him in the transmitter, turn on the scan, and greet him when he walks from the receiver. But is he really the same person? If the system works well, the duplicate will be indistinguishable from the original in any substantial way [117].

There are numerous problems with this description. When the matter transmitter "identifies" an object's atoms or molecules, what is it doing? The term "identity" and its cognates are unfortunately used in several ways in English. Casually, "to identify" can simply refer to determining the kind of thing something is ("we need to identify the mysterious substance found in the victim's car"). Or, it can refer to pointing out an individual ("we will now identify the murderer"). "Identical" can mean simply that two things are of the same type ("the two guns the murderer used were identical models"). But for the issue discussed here, the meaning of "identity" refers to a specific relationship. The relationship of "identity" is a very strong, and necessarily strong, concept that strictly refers to literal sameness — not just similarity. When we use "identical" in this sense, we are not saying that two things are just very like each other ("identical twins"), or even exactly similar ("identical cars"), we are saying that "two" things are actually one and the same specific thing. We may have been using different words at different times to refer to that thing, due to our own ignorance or the specifics of our linguistic needs, but there really is only the one thing. If A is identical to B, then A and B are really the very same, one thing. The classic example here is: "The morning star is identical to the evening star." This statement of identity does not mean that the morning star and the evening star are alike; it means that they are in fact the very same, one thing — one celestial object with two

Why Uploading Will Not Work

names. In the case of personal identity and survival we need this strong sense of identity. If copying my mind is going to save me, or preserve me, then the copied mind must be (strongly, literally) identical to my current mind, not just exactly similar. Otherwise, copying my mind would only be like copying a page in a book — it would produce a new thing exactly similar to my mind but would not be my very same mind. Uploading is going to help us only if copying acts differently with personalities than with physical objects, only if copying maintains this strong sense of identity between the brain-housed mind and the computer-housed mind.

So when Moravec says that the matter transmitter will identify an object's atoms, he presumably is using the term in the casual sense of accurately detecting and scanning and recording information about the particles. But what does he mean when he says "the identity of the atoms would be transmitted to a receiver"? If he means that identity in the strong sense is being transmitted, then this would either beg the question (the claim that copying preserves identity is what we are looking for a defense of in the first place) or it would turn identity into some reified substance that can be moved (going beyond the soul parallel to a kind of even stronger and stranger Platonic realism). Presumably, the claim that the identity of atoms is being transmitted only means that the information about the particles' properties and location is being transmitted and nothing more metaphysically complicated than that. This information is then used to construct a copy of the human being from different atoms in a different place.

But at this point Moravec asserts that the copied person is the same (using the strong sense of identical) as the original, because "the duplicate will be indistinguishable from the original in any substantial way" [118]. This claim however, either begs the question or is simply false — and this is a very important point. The claim would beg the question if the caveat "in any substantial way" means that, in fact, the duplicate can be distinguished from the original, but not in a way relevant to questioning identity — the whole point here is to find out what is relevant to identity, not simply to state that identity is preserved. More directly, the claim would be false if it really does mean that the duplicate and the original cannot be distinguished. They can be distinguished. They are made of different matter, have different causal histories, and are in different spatial locations.

Again, we have to take our terms very seriously here — not for the sake of being picky but because the terms are being used to describe decisive real-world properties. We cannot just use the term "distinguish" in the casual sense of "I personally cannot tell a difference". That just refers to our lack of knowledge and our limited perceptual abilities. Using the term this way, if we could not tell a difference between two twins, then they would really, actually be the same, one person. "Indistinguishable" means literally that there are no properties that one thing has that another thing does not have. If it really is the case then that "two" things cannot be distinguished, they are literally the same thing — and there is only the one thing. Mark Twain and Samuel Clemens are literally indistinguishable — "they" have the same properties. In fact, that means there are not two people at all, only one. This criterion of

P. D. Hopkins

indistinguishability is a good one for determining the strongest, literal relationship of “identity”. It is this criterion that Leibniz used (called the Identity of Indiscernibles, or Leibniz’s law) to determine whether “two” things really were identical [Leibniz, 1902, 14]. In Moravec’s matter transmitter example, however, the duplicate and the original are not truly indiscernible and can be distinguished (whether the original is destroyed or not). They are made of different matter, have different causal histories, and are in different co-temporal spatial locations. Also importantly, it does not help to claim that even though the bodies are distinguishable, the minds are not. In fact, the minds are distinguishable as well — they too are in different co-temporal locations, they are produced by different sets of matter, they have different causal histories, and they will have different futures as a result of all these differences. Neither does it help to claim each mind will think or feel that it is the original — of course they do, but this is because the content of memories and emotions are produced by exactly similar neural architectures. That no more proves that both are correct in their belief than having any memory proves that the event remembered actually happened. False memories are possible — they are not caused by the experience remembered but instead by other neurochemically mediated events such as dreams or illness or brainwashing. In this case, the copy would remember growing up and having specific childhood experiences — but it did not. It came into existence moments ago. Since the minds are distinguishable, one cannot successfully argue that they are identical due to their indistinguishability. Something else would be needed. Moravec goes on to argue that the key is the concept of pattern and pattern-identity.

- Suppose that two receivers respond to the message from one transmitter. Which, if either, of the two duplicates is the real original? The body-identity position on this question is clear. A matter transmitter is an elaborate execution device that kills you and substitutes a clever impostor in your place. The pattern-identity position offers a different perspective. . . After the transfer there will be a copy of me in each one. Surely at least one of them is a mere copy: they cannot both be me, right? Wrong! Rooted in all our past experience is the assumption that one person corresponds to one body. In light of the possibility of separating mind from matter and storing and transmitting it, this simple, natural, and obvious identification becomes confusing and misleading. Consider the message “I am not jelly.” As I type it, it goes from my brain into the keyboard of my computer, through myriads of electronic circuits, and over great amounts of wire. After countless adventures, the message shows up in bunches of books like the one you are holding. How many messages were there? I claim that it is most useful to think there is only one, despite its massive replication. . . The message is the information conveyed, not the medium on which it is encoded. The “pattern” that I claim is the real me has the same properties as this message [118–119].

At least two important things are going on in this passage. One has to do with comparing the mind to information, the other has to do with the nature of patterns.

Why Uploading Will Not Work

As in many other discussions, the analogy used to defend the claim that uploading preserves personal identity is one of information transfer. The idea is that just as we can transfer information (such as computer files) from one place to another while maintaining identity, we can transfer minds. Unfortunately, the analogy does not work. Ironically, the reason the analogy does not work is not because the mind is a special case that operates differently than information — it is because information “transfer” does not work this way either! Look at how the message in the analogy is being treated. The message “goes” from a brain “into” a computer and then travels “over” wire. Just as the more simplistic descriptions of uploading treat the mind, this analogy treats the message as a substance that moves, again employing spatial and motion metaphors. But there is no message that literally travels. There is no thing or object called a “message” that leaves my brain and moves into the computer. My brain causes my muscles to move, which cause my fingers to move in ways that activate my keyboard. The keyboard sends a series of electrons to the computer that organizes other particles in the computer’s storage medium symbolically represented by 1s and 0s. The computer can then send more electrons through wires (or light pulses through fiber optic cables) in a series that can be received by another computer that will be triggered by the organization of the electrons to organize the particles in its storage medium in a specific way (symbolically represented by 1s and 0s). Those particles will then be used by the computer to activate different portions of the monitor to produce patterns of photons which will be emitted by the monitor and enter my eyes, eventually activating certain cells that open channels that release neurochemicals that are then interpreted as the message “I am not jelly.”

While Moravec, and many others, treat “information transfer” as the movement of a thing from one place to the other, using spatial and motion metaphors, it is not the case that a thing called the message “I am not jelly” in fact “moved” from my brain “through” computers and wires “to” another computer and “into” another brain. The sentence on your computer screen “I am not jelly” is not the “same” sentence as the one that started in my brain. It is a physical event that is interpreted by your brain as a sentence. Now, even if someone realizes that in fact no literal object called a message was “sent” “over” wires “to” another person, they may want to say that the information is the same. But this is a slippery use of the term “same” and a tricky use of the concept of information. Information in this view is also being reified and treated as a thing that attaches to symbols or electrons and moves along wires to other places. But what would this extra thing called information be? When we look at the actual physical events that are going on, we can detect channels in neurons opening and chemicals moving around. We can detect electrons moving in wires. We can detect photons moving from monitors to eyes. But where do we detect the additional thing called “information”? We do not. In spite of the linguistic way it is treated, from a materialist point of view, it is unlikely that information is an additional substance that neurochemicals and electrons and photons “carry”. This is a ghostly and soul-like way of thinking about information. Information is what we

P. D. Hopkins

call our interpretations of physical events. The only sense in which the visible physical event of “I am not jelly” showing up on your computer screen “carries” information is metaphorical — you interpret that visible physical event as indicating a state of mind that I have. Despite our metaphors, no ghostly substantial thing called a “message” is being moved anywhere and no ghostly substantial thing called the “real me” or mind is being moved anywhere. Matter and energy is being moved that trigger receiving devices to construct physical systems exactly similar to the originals — exactly similar but not the very same identical systems.

4. The Ontology of Patterns

All this boils down to a confused ontology of patterns. While it is the case (consistent with materialism) that a particular organization of matter can produce a mind, the confusion comes in when we start using terms like “same” to describe another organization of matter and the mind that it produces. The temptation is to draw from the historical influence of soul and ghosts and dualism to reify patterns. Sometimes this is explicit (Moravec says “pattern identity has dualistic implications. . .minds can be separated from bodies [119]), but often it is more subtle. The key passage in Moravec is this one:

- If we were thus to transform a program that simulates a person, the person would remain intact: his mind is the abstract mathematical property that is shared by the old and the transformed programs; it does not depend on the particular form of its program [121].

The central claim here is that a specific person’s mind is an “abstract mathematical property”. If we think of the mind this way — as an abstraction from physical systems that can then be moved around, it is not surprisingly that we can imagine transferring a mind from one place to another. The problem is that this “abstract” property is in fact not being treated as abstract at all. It is being treated as real and it is very doubtful that the mind is in fact this kind of property. The truth behind the “abstraction” idea that does have merit is that the mind is not a supernatural substance that inhabits bodies like vital spirits or souls. The mind is produced by physical systems organized in particular ways. But the mistake — and the way supernatural soul-concepts reassert themselves — is to think of the organization of a physical system like a brain as an abstract but nonetheless reified property that is somehow attached to a physical system. This is the “pattern” that Moravec relies on. But this gives far too much independent reality to a pattern and ends up being inconsistent with materialism.

It is (probably) true that neural activity patterns produce a mind. But the “pattern” is nothing more than the way in which the neurons and associated matter are organized. Unless some form of substance dualism is true, there is no independently existing “pattern” above and beyond the actual matter. Take away the matter

Why Uploading Will Not Work

and you have no pattern left. Patterns are not independent. “They” are just nouns that we use to talk about the fact that material systems are organized in particular ways. To call a mathematical property “abstract” is in fact to acknowledge this. “Abstract” means “existing in thought or as an idea but not having a physical or concrete existence; conceptual” [Oxford English Dictionary], not “immaterial substance or Platonic form”.

Thinking of patterns in this reified way underlies the belief that if you copy the pattern of the brain and reproduce that pattern in another system then you reproduce the mind. The key idea is that the “same” pattern will preserve the “same” mind. But this conclusion is wrong because it equivocates on the word “same”. Yes, it is the case that the matter of neurons organized in a particular way will produce a mind, and the matter of microchips organized in an exactly similar way will produce a mind that has exactly similar hopes, dreams, memories, etc. But this does not mean that they are the same mind in terms of identity. They do not in fact have the “same” pattern. I am not saying here that there are organizational differences between them. I am saying that it is a trick of language to think when one physical system is organized in the “same” way as another that they literally share the “same” pattern. The first sense of “same” just means they are organized in exactly similar ways. The second sense of “same” means the organization itself is a discrete, real thing that is literally located in two different places, inhabiting two different clumps of matter.

Think of it this way. When you print off multiple copies of a document, you do not end up with one document that is multiply instantiated in a (Platonic?) realistic sense. You end up with multiple printouts whose material ink is organized in exactly similar ways. Materially speaking, there is no information on the page. There is ink! The way the ink is organized on one printout, however, is interpreted by our brain as saying the “same” thing as another printout. This is not because there is some reified substance called “information” that is somehow present in all the printouts. It is because “information” is our interpretation of those ink distributions and we interpret exactly similar ink distributions in exactly similar ways.

Patterns (whether associated with ink or neurons, with information or minds) are not ghostly independent realities that can move from here to there or inhabit physical systems. Moravec, like a Platonist, is treating the form as if it is the most real thing and the actual physical systems as secondary. But, to the best of our materialist knowledge, it is the other (Aristotelian) way around. Physical systems are the most real things and the “patterns” they exhibit are just conceptual.

If we come to think that neither souls, nor ghosts, nor patterns, nor minds are moveable independent substances, we can more easily understand some intuitions about thought experiments that highlight the problems of copying personal identity. For example, proponents of destructive uploading say that if your neural patterns are copied to a computer system, and as each computer connection is made your old neuronal connections are destroyed, you have nonetheless survived the process and been transferred to a computer. But imagine this. I tell you that I am going to give

P. D. Hopkins

you a painless lethal injection. However, I also say “do not worry, though, we are copying you right now and only the second that the copy is completed will I inject you. The copy is in fact the very same thing as you so you won’t actually die.” This promise of survival will not comfort you unless you think you will be “leaving” your body and moving into the copy — whether as soul or mind or pattern. Now you may have read about uploading and do in fact believe this, but your position can easily be shown to be problematic because after the copy is made, I could drop the injection and say “just kidding!” and then there is the copy staring at you (perhaps also with a hypodermic-wielding confederate nearby so its experiences are alike). You would not think that the person staring at you is in fact you. But, the fact that in the just kidding case you were not injected makes no difference at all to identity. You are in exactly the same relationship to the copy as you would have been had you been injected. There is no difference. So if you still have your mind in the just kidding case and do not think that copying transferred you from here to there, why think that copying would have transferred you to the other body had you been injected? The destruction of your “old” body is just a smokescreen, obscuring the fact that the exactly psychological similarity between you and the copy is not identity.

Also note that slow and piecemeal destructive uploading will not make a difference either. Destroying your original neurons one at a time while the copy is being made will not solve any problem. Slowly or quickly, the original brain is in the same relationship to the copy. If you are destroyed, you will not be around to ask questions about it, but the death of your brain makes no causal or metaphysical change in the relationship between your brain and the copied brain. Here or not here, nothing changes.

5. Conclusion

Not all transhumanists believe that sheer copying will preserve identity. Kurzweil, for example, challenges the metaphysical efficacy of uploading [2000]. However, there are numerous other examples where proponents of uploading fail to acknowledge or adequately deal with the problem of preservation of personal identity and simply assume that copying will save us. If the criticisms presented here are correct however, uploading may be technically possible but will not accomplish what we want it to accomplish. It will create new minds exactly similar to other minds, but will not save anyone’s life. Actually, I hope that the arguments here are incorrect because I think uploading would be wonderful. However, as things stand, I do not see how a mind can be disembodied and “transferred” and what is more, this is the more thoroughly materialistic and naturalistic position. Uploading sounds at first like a wondrous marvel of technology that promises immortality, but on closer inspection it depends on vague, inaccurate, and faulty assumptions that are holdovers from supernaturalism and dualism. Treating minds and patterns as objective moveable substances or properties is just as mysterious as beliefs about ghosts and souls and vital spirits. The discourse on uploading has inherited a language that tricks us into

Why Uploading Will Not Work

thinking minds work much the same way as souls. The ghosts haunting transhumanism have led us astray.

References

- Alexander, B. [2000] "Don't die, stay pretty," *Wired* 8(1), <http://www.wired.com/wired/archive/8.01/forever.html?pg=3&topic=&topicset>.
- Alexander, B. [2003] *Rapture: How Biotech Became the New Religion* (Basic Books, New York).
- Fact-Archive.com. [2005] "Mind transfer," <http://fact-archive.com/encyclopedia/Minduploading>.
- Human Enhancement and Biopolitics. [2009] "Uploading your mind," <http://hplusbiopolitics.wordpress.com/2009/08/08/uploading-your-min/>.
- Kadmon, A. [2003] "Mind Uploading — An introduction," <http://euolution.com/transtopia/uploading.html>.
- Keim, B. [2009] "Save mind uploading from wikipedia doom," *Wired*, <http://www.wired.com/wiredscience/2009/04/minduploadwiki/>.
- Knowledgerush. [2009] "Mind uploading," <http://www.knowledgerush.com/kr/encyclopedia/Minduploading/>.
- Kurzweil, R. [1999] *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (Viking, New York).
- Kurzweil, R. [2000] "Live forever," *Psychology Today*, <http://www.psychologytoday.com/articles/200001/live-forever>.
- Kurzweil, R. [2005] *The Singularity is Near: When Humans Transcend Biology* (Viking, New York).
- Lakoff, G. and Johnson, M. [1980] *Metaphors We Live By* (University of Chicago Press, Chicago).
- Leibniz, G. W. [1902] *Discourse on Metaphysics* Translated by George Montgomery (Open Court Publishing, La Salle IL).
- Moravec, H. [1988] *Mind Children: The Future of Robot and Human Intelligence* (Harvard University Press, Cambridge MA).
- Moravec, H. [1999] *Robot: Mere Machine to Transcendent Mind* (Oxford University Press, Oxford).
- Sentient Developments. [2009] "Anissimov on the benefits of mind uploading," <http://www.sentientdevelopments.com/2009/01/anissimov-on-benefits-of-mind-uploading.html>.
- Virtual Worldlets Network. [2006] "Neuroprosthetics, brain emulation and mind uploading: The ultimate VR concepts," <http://www.virtualworldlets.net/Resources/Hosted/Resource.php?Name=NeuroprostheticsBrainEmulation>.