John Filiss
Uploading

Retrieved on 29 January 2011 from www.primitivism.com

John Filiss

Uploading

Contents

References

Drexler, K. Eric (1994). "FORUM: Automated police & defense ('Nanarchy')"Extropy #12: 32–39

Hanson, Robin (1994). "If Uploads Come First: The Crack of a Future Dawn" Extropy #13: 10–15

Merkle, Ralph (1993). "Uploading Consciousness" Extropy #11: 5–8

Moravec, Hans (1988). Mind Children. Cambridge, Massachusetts: Harvard University Press

Moravec, Hans (1993). "Pigs in Cyberspace" Extropy #10: 5–7

Penrose, Roger (1990). The Emperor's New Mind. New York, New York: Oxford University Press

Ross, David Justin (1992). "Persons, Programs, and Uploading Consciousness" Extropy #9: 12–16

> "We might then be tempted to replace some of our innermost mental processes with more cyberspace appropriate programs purchased from the AIs, and so, bit by bit, transform ourselves into something much like them. Ultimately our thinking procedures could be totally liberated from any traces of our original body, indeed of any body. But the bodiless mind that results, wonderful though it may be in its clarity of thought and breadth of understanding, could in no sense be considered any longer human." (Moravec 1993: 7)

So much, incidentally, for pattern-identity.

At some level, uploading may be the most perfect (if unintentional) method of wiping out the human race ever devised. An individual is destroyed and replaced with a reasonable facsimile, the continuity between the two established through reductionist arguments which define away the uniqueness of individual consciousness.

But that is part of a more distant future which is difficult to predict. In the present and near future, perhaps the most dangerous aspect of uploading theory is the saccharine gloss it lends to many aspects of our cybernetic future. As our lives become more deeply enmeshed in a technocratic web, then the prospect of merging with the machines which we will be increasingly subject to is tempting. If our future forebodes an eclipse of the organic by the automated, then perhaps the most hopeful response is to embrace the coming wave.

Looking back over the concepts and theories underlying the prospective science of uploading, it recalls nothing so much as the sarcophagi and totems of past civilizations, an attempt to inscribe an eternal imprint of oneself in the ceaseless void, a desire to forever be. However noble and transcendent such a vision might appear, it will have to be based on some semblance of critical rather than wishful thinking, lest it become a tool of our future enslavement, as it has been of our past.

Whatever humanity's potential for immortality, it will have to do better than this.

Uploading (occasionally referred to as downloading) is the projected science of transferring human consciousness and memory from organic tissue to an automated facsimile, usually described within the narrower confines of transferring mental functions to computer.

Little-known outside of technophile circles, uploading remains the most controversial of all possible technologies wherever it is discussed. This, in spite of the fact that seemingly no speculative technology in history — not even nanotechnology, by itself — can make greater claims to granting extraordinary powers to humanity.

The range of possibilities open to a conscious being in cyberspace is difficult to even begin to visualize. You would become effectively immortal within a computer program, immune to disease, aging, or injury. You could inhabit a fantasy world not subject to our physical laws, possessing the power to metamorphose into any form, or instill your consciousness into any object within the program. You could possess the power of flight, or the ability to perform telekinesis. You could modify your existing environment on a whim into forms unknown on Earth, or seemingly anywhere.

Not that an upload is limited to the environment within a program. If the concept is feasible, then we should be able to place the computer program within any vessel that can sustain it. There is already some interesting speculation on this score.

Both utility fog and Moravec's robot bush (Moravec 1988: 102–108) would be possible contenders for receiving an upload. Less empowering, perhaps, are the speculations of Robin Hanson. He foresees miniature uploads functioning at high speed:

> "Faster uploads who want physical bodies that can keep up with their faster brains might use proportionally smaller bodies . . . .a 7 mm. tall human-shaped body could have a brain that fits in its brain cavity, keeps up with its 260 times faster body motions, and consumes 16W of power. Such uploads would glow like Tinkerbell in air, or might live underwater to keep cool.
>
> "Billions of such uploads could live and work in a single high-rise building, with roomy accommodations for all, if enough power and cooling were available. To avoid alienation, many uploads

might find comfort by living among tiny, familiar-looking trees, houses, etc., and living under an artificial sun that rises and sets 260 times a day. Other uploads may reject the familiar and aggressively explore the new possibilities." (Hanson 1994: 11)

Before delving too deeply into the controversies surrounding and the theories underlying uploading, some idea of what the uploading process might consist of should give the reader a better feel of what is actually being attempted.

One elaborate scenario (Moravec 1988: 109–110) involves a robot brain surgeon who opens a human patient's anesthetized skull, and places its hand on the brain surface. The hand is bristling with microscopic instrumentation, that can scan into the first few millimeters of brain surface. High-resolution magnetic resonance measurements build a three-dimensional chemical map, while an assortment of electrical antennae register the pulses flashing among the neurons. A computer attached to the robot stores the above information as a program based on the scanned brain tissue. The patient is furnished a push-button that allows him/her to test the stimulation. When it is pressed, electrodes in the robot's hands are activated that override the normal signaling activity of the scanned neurons. For as long as the button is pushed, a small portion of the patient's brain is replaced with a computer simulation. After pressing the button enough times to be certain that there is no difference, the patient allows the simulation to be activated permanently. The scanned brain tissue is now impotent — it sends and receives signals as before, but its output is ignored by the remainder of the brain.

Microscopic manipulators on the robot's hands carefully excavate the now superfluous brain tissue, and vacuum it away.

The robot's hand sinks slightly deeper into the brain, and the process begins anew. Eventually the skull is empty, with the robot's hand resting deep in the patient's brain stem. The mind has presumably been transferred from the human's body to a computer. In a final, dramatic act, the robot lifts its hand from the skull. The connection broken, the body shudders and dies.

Better to be alive for but one day while retaining the capacity for joy, than to exist for an eternity as a feelingless processor of data.

But, leaving even this aside, what might our future be in a post-upload world? Despite our newfound powers and abilities, it shouldn't be long before we are upstaged by AIs unburdened with the need to carry a human upload. These AIs should be far more efficient and adaptable than uploads, and may well surpass us at some point in every endeavor. Moravec, much to his credit, is one proponent of uploading with the honesty to approach this issue.

"A human would likely fare poorly in such a cyberspace. Unlike the streamlined artificial intelligences that zip about, making discoveries and deals, reconfiguring themselves to efficiently handle the data that constitutes their interactions, a human mind would lumber about in a massively inappropriate body simulation, analogous to someone in a deep diving suit plodding along among a troupe of acrobatic dolphins." (Moravec 1993: 7)

With projected future gains in computing power, the expected advent of nanotechnology, and a more sophisticated approach, the future for AI looks good. If Eric Drexler, originator of nanotechnology, is correct that post-nanotechnology machine intelligences can perform a million years of research and development in a calendar year (Drexler 1994: 35), then it seems we could be outclassed in a very brief period of time by AIs with a code of values we can only speculate on. Though some or most AIs may be non-hostile, the economic determinacy of a world facing such blinding technological progression, with its concomitant extreme competition, may make them limit our sphere of movement more and more so as not to hinder their own development.

If we can not perform any useful function as a primitive holdover from a carbon-based life-form, then our importance on the frontier of scientific development should diminish altogether. When we reach the point where we are no longer useful, we should perhaps prove a liability to those Al's most well disposed towards us.

What might be the response of an upload that faced successful competition from more efficient Al systems? Moravec writes:

design and construction of newer, smarter, and more powerful machines. The projected trajectory of a post-upload future is in some sense dependent on both these abilities.

The prospect of making a computer copy of oneself likewise seems well within the bounds of the possible. It may be more difficult than proponents of uploading anticipate. It is difficult to know for certain if even an upload comprising a simulation of each neuron, synapse, and nerve impulse traveling along every neuron in a human brain would in any way behave like the original. Memory storage in the human brain remains something of a mystery, and there may be unforeseen difficulties in giving a copy the memories of the original. But overall, copies seem not impossible.

Transferring consciousness from a human mind to a computer is, however, another matter entirely. There is not one solid argument nor bit of evidence as to why it should be possible. The theory underlying uploading — pattern-identity — is without foundation: more than anything, it seems like an extension of humanity's unfortunate habit of confusing an existent with its representation. The map is not the territory. Patterns may well be lifted and replicated, but identity remains unique and inviolate.

Another, little-discussed difficulty springs forth at the suggestion that the self can be transferred from body to machine. Could the individual's emotional life be uploaded onto a computer? What would the emotional range of a neural upload be? The argument that we could transfer our emotions from a human body to a machine seems an impossibility of a higher order than transferring mental function. Simulating certain mental functions is, after all, what computers were designed to do from their inception. But emotions seem a different problem entirely.

Though mental processes may, by and large, be relegated strictly to the brain, emotions seem to well up within and throughout the body. They involve the breathing, the heartrate, the endocrine system, the musculature, as well as the brain. Would a neural upload experience anything like a human's capacity to feel? Would it experience anything at all?

It is important to point out that emotions, virtually ignored in all discussions of uploading, are essentially what give our lives meaning.

Another scenario (Ross 1992: 16) involves injecting nanomachines into the bloodstream that would replace each brain and sensory neuron with a functionally equivalent, artificial structure. The nanomachine would contain a program that would emulate the neuron, while at the same time interacting with neighboring cells as though the replaced neuron were still in place. The cells surrounding the neuron would be unaware of any change. Gradually, each synapse in the brain would become information in a computer program, retaining functionality but dispensing with its former physical structure.

When the process is complete, what is thought to be the individual would awaken to a new life in cyberspace.

Pattern Identity

One occasional misunderstanding that arises in discussions of uploading is the difference between a copy and a transfer. The question of whether the latter is even possible is the fiercest and most fundamental controversy surrounding uploading.

A copy is a simulation of an individual which may be similar in many, many ways to the original, but which does not purport to be the original. Given the current rate of technological progress, the ability to make copies at some level, including ones that could pass the Turing test, or fool friends and relatives, etc., seems possible.

A transfer is much more difficult, if in fact doable at all. Transferring consciousness from brain to computer implies copying at a deep enough level, at least to the extent of replicating individual neurons, while destroying the original and maintaining the functional integrity of the whole during the process. Such are, at least, the very minimum constraints that can be presumed for any foreseeable uploading process, but they fail to answer the most crucial question: What evidence or line of reasoning exists that uploading can actually be done?

The theory put forth by proponents of uploading to support their belief in its viability is called pattern-identity. Moravec offers the most complete exposition of this viewpoint in relation to uploading

in print, and the following is a summing up of and response to his arguments (Moravec 1988: 116–122).

Pattern identity rests on the basic premise that the continuum of life is defined by pattern and process, not the substance that supports it. Moravec counterpoises the pattern identity position with what he calls body-identity, the idea that an individual is defined by the substance with which he or she is made. Though some interesting arguments are mustered in support of pattern identity, one need not rigidly adhere to a body-identity position to note their weak points.

He begins by observing that the preservation of pattern and loss of substance is a normal part of organic life, that humans eat and excrete, old cells die to be replaced by new, parts within the cell are slowly being rebuilt and replaced, etc.. His strongest argument, it hints at some measure of truth to pattern-identity, and bears further examination.

Though it is true that substance does shift over the course of life functions, so too does pattern. An individual might, say, lose a thousand neurons from drug use, and obviously remain the same person, though the pattern of neurons has changed. The effect may not even be noticeable. Actually, patterns that can be derived from our organic states shift constantly without affecting identity. We all change over the course of time, yet remain ourselves.

The reader may by now notice the basic branch of science that is ignored in uploading theory: chemistry. In chemistry, the medium is the message, as it is in direct perception. Different substance, different pattern.

Though a given message or pattern may be conveyed by different media, any illusions of identity end there. The media (in this case, organic neurons vis-a-vis mechanized computers) obviously possess different properties and a different "life."

Pattern-identity as described by Moravec is, at best, a one-dimensional description of life, not a wondrous key to its furtherance.

It should also be noted that the very gradual changes in substance seen in metabolic processes are under the complete control of the organism and not the result of some outside force, such as busily working nanomachines, acting upon it. Obviously, all changes in

wealth to machines. No longer would fortunes, great or small, be passed on to the next generation, but would instead "remain" in the control of individuals uploaded onto a computer program, or at least computer programs which can give a convincing simulation of individual personalities.

Another effect would be to make a growing portion of the labor force subject to the authority of machine employers, as not only personal wealth would be retained by the upload, but presumably valuable professional skills and experience. Since uploads are not expected to require sleep, and may conceivably operate at speeds thousands of time faster than a human being (an upload with sufficient processing speed could perform, say, a month's worth of research and project analysis in the space of a lunch break), they would prove enormously valuable, and eventually indispensable, to the operations of major business, legal, investment, and consulting firms, as well as in universities, think-tanks, and scientific research. In fact, due to the much higher speeds and probable efficiency gains, an individual could be worth far more as an upload than as a flesh and blood human.

The Coming Wave

The implications of uploading are so vast that it is difficult to sum up the many consequences and controversies without invoking new ones. To begin with, the proposition that uploading is both possible and desirable rests on a whole string of assertions. Let's take a view of each.

One very basic issue is whether computers can ever be capable of thought, or become self-evolving. Though machinery is comprised of different substances with different structures and functions than that of organic life, there is no reason to believe that computers will be incapable of powers which would not fulfill even the most demanding definition of intelligence. Similarly, it is perhaps possible to imagine a future where machines are wholly responsible for the

Ross sees intelligence as more an issue of the system's "complexity" (ibid: 12–13), a correct, if rather vague assertion. The defining aspects of intelligence, its flexibility and fluidity, its ability to engage in a bewildering variety of tasks, can only be simulated or achieved by the complex interaction of a vast number of individual subsystems, none of which need possess extraordinary ability, nor even be crucial to the system's functioning.

Though it is hard to fault such a definition of intelligence as far as it goes, it is as evasive as it is explanatory. It may be sufficient to explain the means by which we can create AIs that can pass our subjective tests of intelligence, but it doesn't really explain exactly what intelligence is. And the impossibility of so far doing this makes inconclusive the argument that computers can't think. If it is impossible to define exactly what types of interactions give rise to intelligence in humans, it is likewise impossible to claim that intelligence can never appear in machines. All that can be said for certain is that humans are not machines, nor vice versa. Humans and machines may share many similar attributes, but not identity.

Another consideration that could affect the expected time-frame for uploads is the issue of how deep a level must be replicated in order to support an uploaded consciousness. Must we simulate each individual atom, or perhaps each molecule, or possibly every subcellular organelle, or, optimally, can we get away with merely simulating each neuron? The difference in the amount of computer memory needed is substantial. Merkle estimates that sometime between 2010 and 2020, the amount of memory needed to store an atom by atom description of the brain would occupy a volume of somewhat over 100 liters. In the same future time period, a computer that would simulate each brain neuron, synapse, and nerve impulse would occupy a space of only one cubic centimeter (Merkle 1993: 5,8). As even the possibility of uploading is conjectural, the degree of simulation needed is uncertain.

Another important consideration which would affect the adoption of uploading is the social environment. It is too early to predict how accepted the concept of uploading and its promises might become, but some of the social consequences of a world with uploads can be conjectured. One effect would be a gradual loss of the world's

pattern and substance must occur within a very narrow framework for life to be sustained.

In addition, there is no available evidence of life forming from other than carbon-based molecules anywhere on Earth, or to our knowledge, the Universe, despite billions of years for it to have occurred. This means that not only did no non-carbon-based life evolve by itself, but no carbon-based life-form has ever shifted its chemistry to a wholly new set of elements, despite what must be overwhelming evolutionary pressures to do so (in order to exploit new substances, new properties, and new environments).

Further argument by Moravec in support of pattern-identity shows well the reductionism inherent in the theory.

Moravec begins by stating the message "I am not jelly." (Wishful thinking, perhaps, for those unhappy with their existence as protoplasm.)

> "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 it is most useful to think there is only one, despite its massive replication. If I repeat it here: "I am not jelly," there is still only one message . . . .The message is the information conveyed, not the medium on which it is encoded. The "pattern" I claim is the real me has the same properties as this message." (Moravec 1988: 118–119)

The confusion here lies between an existent and its symbolic representation. A symbol is, by definition, something that stands for something else, a minimalist rendering of reality that can be reproduced ceaselessly. In the example that Moravec gives, what was real and unique was the whole mental process which made Moravec come up with writing "I am not jelly." The rest is mere representation.

Confirming Transfer

If uploading theory is questionable, what then of the evidence? When the technology is available, couldn't we experiment with different processes and see what the results might be?

Some problems arise here. To begin with, since uploading demands that the original brain tissue be destroyed, or at least rendered inert, the expected death of the patient should make obtaining volunteers difficult, not to mention how the courts might view the matter. This difficulty can be bypassed by waiting, as Ralph Merkle suggests, until the prospective upload is dead (Merkle 1993: 5).

Assuming this condition can be fulfilled, and it is possible to quickly map and/or preserve the neurons from deterioration (which shouldn't be difficult with nanotechnology; upon the cessation of vital signs, subcutaneous repositories with sensors could release hordes of self-replicating nanomachines into the bloodstream to preserve and protect the neurons), another problem apparently arises. Being already dead prevents what some consider the most reassuring indicator that it is in fact you making the transfer.

> ". . . if the person is not conscious . . . there is no way for a person looking forward to such a procedure to be sure he would survive." (Ross 1992: 15)

Actually, the idea that being awake through the "transfer" somehow confers certainty that the whole operation is performing seamlessly is totally groundless. During an upload, the neurons of the brain, the seat of consciousness, are replaced with nanomachine actuators that interact with neighboring cells as though the replaced neuron was still present. Your brain is being replaced, bit by bit on a microscopic level, by machines sophisticated enough to fool the untouched cells that nothing untoward is going on. Under these circumstances, what indication would there be that the transfer is not taking place? Sensory input would be synchronized between nanomachine and neuron, so the world around you would appear the same to all senses. Would some sense of instinctual angst or malaise perhaps rush over you during the uploading process? There

is no reason to believe something like this would occur, nor would it indicate that you are being killed rather than merely shifted from one vessel to another.

More bizarre is the speculation that the copy itself would be aware that it is not the original. In his article on uploading in Extropy, Ralph Merkle casts a fictional scientist asking a computer upload, "Do you think you're not you?" to which the ready response is "Nope, I'm me." (Merkle 1993: 5) Even a relatively simple computer program should be able to avoid a response that is a clear contradiction in terms, e.g., "I'm not me." A less nonsensical question would be to ask by name, e.g., "Are you John Smith?", though this still assumes that a computer program would have an identity other than its programming. Actually, such a question could be useful for a copy, in order to see if memory was actually read in, but confirmation of a transfer continues to elude us.

Future Prospects

It is difficult to predict the future course of a science at once so conjectural and so controversial, but it is likely that something resembling an uploading process could be attempted shortly after the advent of nanotechnology. There has been an attempt by some theorists, notably Moravec, to estimate the needed computer speed and memory for sustaining an upload, then plot the point on a time-line graph to predict when such machines will be available. He estimates that a 10 teraops (10 trillion-operations-per-second) computer with 10 trillion words of memory would be sufficient, and he predicts such computers will be both available and affordable in the year 2030 (Moravec 1988: 59–60, 68).

In one of the more thoughtful essays written on uploading, Dave Ross points out that computer processing speed does not necessarily correlate with program intelligence (Ross 1992: 12). An intelligent computer program, whether it be an upload, copy, or some form of artificial intelligence (AI), could, with sufficient memory, run on the simplest of computers. It would simply be correspondingly slower.