[Existential Risk / Opportunity] Singularity Management

April 2016

Contents:

The "Uploading" Branch of Singularity Thought

by James Blodgett

This publication is about singularities. One part of the attraction of singularities for many people is the concept of "uploading." In order to be up to speed when thinking about singularities and understanding some of their proponents, it is important to know about this concept.

The idea is that exponentially increasing technology may allow us to measure the state and the interconnections of all of the neurons and synapses in our brain. This might be accomplished in several ways. Something like MRI brain scans might increase in resolution, brains might be invaded by swarms of nanobot probes, or preserved brains might be sliced and scanned at atomic resolution. Then the resulting "connectome" might be simulated in a computer. The hope is that such a simulation would reconstruct our consciousness, "uploading" our mind into the computer. Once there, we could experience a simulated environment, or interface with the real world via video or via control of a robot body. In such a state, we could live as long as the computers could be kept running, a result somewhat like going to heaven. Some religions expect an end times "rapture," in which all of the faithful will go to heaven. The upload version of the singularity has been called "the rapture for nerds."

Frankly, I am somewhat dubious about whether this is going to work. I question whether we will ever be able to measure the state and the interconnection of all of our trillions of synapses to adequate resolution. My personal interest in singularities is driven by the good they could do for humanity in general, not by the expectation that they will save me personally by this method. However, recent advances in brain preservation

suggest that adequate resolution just might be possible. But I, somewhat agnostic on this issue, am not the best person to present it to readers. Therefore I introduce readers to Jeff Bone, a knowledgeable advocate of cryonics, i.e. of freezing the body and/or head after death in the hope that advancing technology may permit resurrection. He is not currently a member of our SIG; I met him when I worked with him on a Lifeboat Foundation grant proposal. He is a serial entrepreneur and a good-guy hacker of computers, of personal life, and of human futures who has been involved in many interesting projects. His interest in cryonics is directly relevant to a discussion of uploads because he thinks that a likely method of resurrection may be an upload procedure, so he is also knowledgeable about that.

Interview with Jeff Bone

Interview of Jeff Bone by James Blodgett

Blodgett: Jeff, I learned of your interest in cryonics and uploads from an old email. In that email, you advocated cryonics as insurance valid unless the future contains some other form of resurrection. You mentioned Tipler's ideas about some advanced technology recreating us in the far future. Of course some people hope for a spiritual resurrection. Cryonics was available in Heinlein's day. Since he was a great science fiction writer who was comfortable exploring many futures, a cryonics advocate asked him why he had not signed up. He answered that he wasn't sure about its affect on the afterlife. I suppose St. Peter might have to wait a long time if we are playing in simulo, but it would seem that he could wait until the computers stop working due to heat death of the universe in a few trillion years. We would only miss a few trillion years in the real heaven. Bostrom postulates that our era will be fodder for historical simulations. Indeed if there are many such, and if we are important enough to simulate, there will be many versions of us, and since a simulation that takes seriously the reality of our current choices and therefore makes those choices realistically doesn't know it is a simulation, and since (given these conjectures) there will be many versions of us, the probability is high (given these conjectures) that the version right now reading this is one of the simulations. Does the continuity you hope for lose its value if multiple?

Bone: Not at all. Indeed I would say that continuity is a bit of an illusion anyway; our own physical (or, at any rate, current) persons are interrupted and discontinuous in everyday situations — brain activity and dreaming during sleep are clearly internal process discontinuous with our everyday experience. And in extreme circumstances such as brain trauma or other insult we often see the cessation of normal brain activity only to have it restored later.

Given these observations, we have to ask about the nature of this perceived continuity that we seem to cherish. I would say: it's a *perceived* and consistent history that gives us our identity. Hence any historical "thread" can be perceived as continuity of identity;

all copies and any substrate are therefore ontologically equal. Each would think — and have a valid claim to — being "me" in some sense.

All that said, in a case where you might have multiple overlapping instances in the same substrate / perceived reality, it would be a great boon if "merging" of divergent copies / histories of yourself was an eventual possibility.

Incidentally, as an additional item on the plus side of uploading, it might be worthwhile to mention that there is *some* possibility, assuming the cosmological constant is sufficient to support a closed rather than open universe, that future intelligence could be able to harness and manipulate the collapse, or use some other quirks of physics, to avoid the heat death and provide subjective infinite time.

Blodgett: What is your estimate of the probability that cryonic resurrection will work? What method do you think most likely? What is your strategy and philosophy about the odds?

Bone: I think cryonics is a long shot now; however, it's something that *could* physically work, whereas most other alternatives (i.e., simply hoping / believing in an afterlife) are rather epistemologically ungrounded. And even within cryonics as a whole today, different approaches give different probabilities of a positive outcome. Unfortunately it has been the case that the methods used to date generally create substantial structural damage at the cell level within the preserved brain.

My own assumption / goal is that the brain is the seat of consciousness, and it's the dynamic electrochemical activity of the brain that gives rise to the transient consciousness while the structure encodes memory, experience, identity and so forth. So an appropriate simulation of structure, coupled with a general dynamics, would be equivalent to biological personhood.

Anything that can preserve the structure with high enough fidelity should, in effect, allow for "resurrection." So preserving the structure without damage is key. This is the motivation, for example, for the "head-only" preservation methods — cooling and preserving only the head can be done more quickly with less damage to the neural tissues. There has been substantial progress just in the last few years in vitrification, perfusion technique, and other improved structure-preserving methods. I believe that these, in combination with high-quality post-mortem brain scans and better understanding of the electrochemical dynamics of the brain will, eventually, lead us to being able to "restore" a biologically deceased person.

Blodgett: What do you think about our potential life as an upload? To what extent would it really be us?

Bone: I believe that the question is a bit open — to what extent am "I" the same person "I" was ten years ago? I think an upload would be just as much "me" as the person answering this question right now. However, I also think the nature of the upload environment needs to be considered as well. Today, I have a familiar sensorium — my set of senses shows me the world around me and enables my interactions with the world itself. When I'm taken out of that sensorium — for example, in sensory deprivation experiments — my brain reacts in odd, perhaps unhealthy ways. And my sensorium today implicitly includes a specific, slowly-changing embodiment, proprioceptive context, and so on. It's likely that we're going to need something equivalent to that as an upload; so one interesting question is, to what extent will we be satisfied with embodiment within virtual environments versus wanting / needing some sort of embodiment in the physical world? At some point, I believe both will be possible; and generally, I believe that those whose brain structures are uploaded at some point may, over an extended lifespan, want to live in both biological and cybernetic bodies, in simulo within high-fidelity virtual environments — or, some case, perhaps both concurrently.

One concern / fear is that uploads could be second-class citizens, easily exploited and enslaved to perform mundane tasks of automation where human-level intelligence is required. This is, probably, the biggest potential pitfall. My own hope is that family / descendants will advocate against this in order to enjoy the company and full participation of their antecedents within their own lives.

Blodgett: I have already seen low- resolution resurrections. I have seen actors recreate historic personages who were prolific writers. If the actors study their subject carefully, they can assimilate some percent of their thought, perhaps a fairly large percent of their important thought. Artificial intelligence could do the same thing. An AI learning to play me may read this text. Good luck with playing me. I don't see it as really me, but in a loose sense it recreates some of me. Bostrom postulates that an intelligent computer may or may not care about personal survival, but that it will want to achieve its objectives, so it would care about (or at least work for) preservation of its objective function, in itself or in other computers that can carry on its work. Humans can see value in that too. A writer achieves a form of immortality as long as his writing is read. I would rather see the future in person or as a high-resolution simulation, but if my ideas about appropriate management of singularities survive, in my writing or in a low-resolution simulation, that may do some good that I would like to see happen. What is your take on this lower level of resurrection?

Bone: I tend to see this as a qualitatively (not to mention quantitatively) different thing. In particular, this kind of resurrection is static; it's based on an incomplete sampling of externally-expressed brain state that is halted, frozen in time if you will, at the moment of death. In my opinion, you have to have as complete as possible a snapshot of brain state,

at the time of “death,” and the ability to run the dynamic processes of the brain going forward over that snapshot, in order to make any kind of claim of continuity of identity.

That said, this is a deep, ontological, epistemological question. There’s no doubt that there is value in even low-fidelity forms of preservation of thought, intent, and so on.

Blodgett: You assume that the form of low resolution emulation of a person I mentioned, i.e. a re-creation of a person based on writings etc. rather than on brain scans, would be static. That is not necessarily true.

I see why it would be true in some versions. It would be true if we insist that the emulation be true to what we know of the original's work, a "snapshot" as you say, and that it not try to develop related thoughts any further. We end up with an entity somewhat like Disney's audio-animatronic Lincoln, condemned to eternally recite a speech derived from Lincoln's writings with no deviations from the original text. However, the degree that such a thing is static rather than dynamic is a design consideration. The Greeks "resurrected" Socrates in the literary form of Socratic dialogs. Some of those dialogs may have been an attempt at transcription of his actually expressed philosophy, but others seem the result of creative license, exploring what Socrates might have thought about a novel topic. Our emulations would seem most interesting if also given such license.

Bone: That’s an excellent point; and, fair enough. What I really meant to say is this: if the emulation does not encompass a fine enough and comprehensive enough “internal narrative” / memory of the subject’s past — an internal history, if you will — and a dynamics for extending that, then I would say it’s not really the same person or kind of thing. It’s certainly within the realm of possibility to simply emulate based on externally-observed / recorded experiences; and if the dynamic simulation going forward perceives the mass of input as, in some sense, its own memory then yes, I’d say you’ve got — something. <grin> I’m not sure what, but we have not historically had the need to make any sort of fine gradations of what is and is not personhood. In a future with a range of simulation / emulation types and embodiments I would expect at least a pop folksonomy, legal distinctions, prejudices, etc. to arise around such distinctions.

Blodgett: Do you recommend cryonics to readers?

Bone: For me, it’s a kind of reverse Pascal’s Gamble. It’s a bet based on extrapolation of observable, physical reality. Other “hopes” for any sort of afterlife are in contrast based purely on faith; it’s impossible to handicap them in any sort of objective way. Even if the finite chance of resurrection through cryonics is vanishingly small, it is still — for me — qualitatively and quantitatively infinitely more likely than other more traditional notions of an “afterlife.”

Again, though, I think this is a highly personal choice, dependent on many subjective factors. Even as an outside shot it seems to me that the downsides are few; but that's based on a lifetime of accumulated experience, thought, and bias on my part. Other people will certainly weigh the factors differently given their own different lives — Heinlein being a prime example.

Blodgett: Uploads are a recurring theme in science fiction. What are your favorite science fiction depictions, and what do they tell us?

Heh, this is the hardest question you've asked! ;-) I'm really a voracious reader and book lover, and science fiction in particular has always provided a kind of "structured" way to explore various ideas. On this topic there's quite a bit of stuff, of course, and of varying quality. However I suppose that the following might be considered a short list of things I've enjoyed; it includes some titles not only on uploading but also cryonics and the transhuman and posthuman condition. These represent merely the tip of the iceberg...

The First Immortal by James Halperin
Tech Heaven by Linda Nagata
Down and Out in the Magic Kingdom by Cory Doctorow
Accelerando by Charles Stross
Everyone in Silico by Jim Munroe
Mindscan by Robert J. Sawyer
Feersum Endjinn by Iain M. Banks

All of Greg Egan's stuff, but particularly Permutation City.
All of Hannu Rajaniemi's stuff

Non-fiction

The Physics of Immortality by Frank Tipler
All of Tielhard de Chardin's stuff

Nb., it was Tipler's aforementioned book that caused me to think differently about the continuity and nature of identity.

Blodgett: Thank you for sharing this with us.