The JOURNAL *of* GEOETHICAL NANOTECHNOLOGY

Terasem Movement, Inc.

Volume 2, Issue 1
1st Quarter, 2007

Terasem Movement, Inc.
201 Oak Street
Melbourne Beach, Fl 32951

Editor-in-Chief: Martine Rothblatt, Ph.D., J.D.
Managing Editor: Loraine J. Rhodes

The JOURNAL of GEOETHICAL NANOTECHNOLOGY

Volume 2, Issue 1
1st Quarter, 2007

Marshall an author, public speaker and founder of HowStuffWorks, offers an understanding and explanation of the pace of technology change how he believes in two to three decades, 'mind uploading' will work.

I might not actually have free will. When I raise my arm, I might not have actually done that; part of my subconscious might have done that and caused me to raise it. I perceive it is free will but it may not be.

Even though I have perceived there is one me and I think of myself as a single person, there might actually be multiple things behind me that are being integrated into an illusion that I am me.

Now I'm supposed to talk knowing that I'm hallucinating an illusion and whatever else. I have to put all that aside and go back to my normal mode of thinking which is: I am one person with one consciousness. I do have free will and I am not hallucinating.

I talk to a lot of people and do a lot of stuff that is fun in the way of educating people, such as with the website: How Stuff Works. One thing I know about talking with the general public is that no one is thinking at the level that is being thought of here and no one is sitting around in their living rooms watching television and thinking, wow, in 20 or 30 years, I can have my brain uploaded. That is just not in the public consciousness.

I have to work at a little bit different level when trying to help people understand the pace of technological change. To help people understand the pace of technological change, I can't use computers because most people don't have a real good grasp of computers. I can use airplanes because everybody understands airplanes.

If you look back to 1903, and at the moment this happened in normal society, there were no skyscrapers and there were not cars yet because the model-T was not invented until 1909.

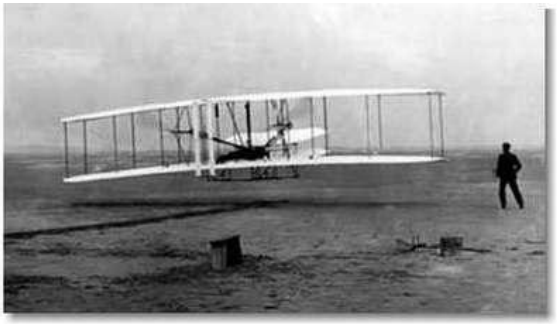

Image 1: Wright Brothers

There wasn't air conditioning, refrigeration, lighting was still - some of it was electrified but a lot of it was kerosene. The concept of the galaxy had not been invented yet so if people looked at the stars, no one thought of galaxies yet because that does not get invented until 1920.

This rickety, wooden, fabric thing takes off, off the ground, and flies for 200 feet. And if you were to say to people in 1903, hey, we just had the first airplane, now think about this, 50 years from now, there's going to be a giant aluminum version of this, except that it's going to be about three football fields long and it's going to be able to fly faster than the speed of sound and it's going to be able to carry 70,000 pounds of bombs around, all the way to the other side of the world, and drop them on foreign nations if it wants to. That will all happen in 50 years.

They would have just thought you were nuts and yet 50 years later the B-52 bomber, which is able to fly halfway around the world and drop 70,000 pounds of bombs on people, actually happened.

Image 2: B-52 Bomber

Now, 15 years, and not 50, is the pace of technological change. That is phenomenal and as Ray Kurzweil suggests, the pace is accelerating. Paradigms are shifting at a faster and faster rate.

It is hard to predict the future, but one thing that I'm pretty sure of, and something that I try to talk to people about, is that we all are going to want out of our human bodies.

We can actually look at market forces that will drive us out of our bodies and understand that is a way of understanding a little bit about uploading and what will drive that.

What will drive us out of our bodies? I can tell you one thing is travel. Travel can be inconvenient: flight connections, security probes, and all the things involved with traveling. The experience of travel is one thing that will drive us out of their bodies but for a lot of people another thing that will drive us out of our bodies is video games, the desire to experience video games much more intimately that we do today.

Here is an image of video game technology in 1982 versus video game technology in 2005. In 25 years we went from Pac-Man, which is four colors on a black screen, to an immersive 3-D environment.

Image 4: Video Games

What would this screen look like in 2030? If we went that far in 25 years, what will a video game look like in 25 more years? And that's mind boggling. What will I look like?

If we stay inside our bodies, it will not look that much different. It can get a little higher resolution but you can get much better than HALF Life II in terms of resolution.

They offer us all these different experiences that we would all like to have, but look at some of these experiences, you can play football,

you can kill people in realistic battle situations and you can go back to ancient Rome.

Image 5: Immersive Video Games

There are lots of cool things you can do but the problem is you have to do it with two thumbs and most of us don't want to experience football and Ancient Rome with two thumbs. We want to experience them with our full complement of physical senses and muscles. We want to actually participate in these events.

This is a rapidly evolving technology. I believe it is either on the cusp of or already has overtaken the movie industry. It has nowhere to go but forward except for this problem: the notion that you are going to control and experience with your thumbs is nuts. That is one thing that will drive us out of our bodies. And it will drive us out of our bodies in one of two ways. Either we will install hardware that will let us emulate or connect into these virtual environments and control and feel them or we'll realize we don't need our bodies anymore. One way or the other will get us out.

The second reason is porn. We all see the effects of porn in our society and I can offer you an interesting and sometimes shocking piece of data to show how popular porn is. I would say we all use Google. Google represents 2.7 percent of all Web traffic, followed by Yahoo! and MSN. Search is about five percent of network Internet traffic and we all use that.

Porn is over three times more Internet traffic than all of search. It's insane how popular porn is and we don't realize it, maybe because it's not something people talk about. But that statistic cannot be denied. That is an astounding statistic.

The way people experience porn right now is through still images or grainy videos and that stinks. It's just not how people want to experience porn. It is a very, very poor stimulation of what people want out of porn. That's the second thing that will drive us out of our bodies.

The third thing is this horrible problem our bodies create with longevity. I was on this flight, wishing I wasn't inside my body, and I'm sitting in my seat, which is the aisle seat. There is a quite large woman sitting next to me who got there before me and just put up the arm rest. We're squeezed in a two pack on the airplane. She's about 80 years old and she used to work at IBM. I got to know her very well on this flight. She lives near Orlando, but is flying up to visit her niece who lives in Burlington, VT. She said to me at the end of the flight, "I really appreciate you being here today." And I said, "Why is that?" And she replied, "Well, my husband of 53 years passed away and this is the first trip I've ever taken without him."

What do you say to that? She talked through the whole flight and I talked with her and we had a very nice time, but I didn't realize I was taking the place of her now deceased husband, That's just shocking to think that I was in that role, for one thing. And for another thing, think how horrible that is, a person she's been with for 53 years just vanished out from

underneath her for no reason. Death is an insult; it's just ridiculous.

There are all these different ways for us to die. We could go to the bathroom and fall down the steps or we could get into an accident as we're driving back to a hotel or the hotel could burn down while we're sleeping; there are so many ways for us to die.

When we die we don't have a backup system or anything like that. Getting out of our bodies is one way to improve longevity.

All of these forces, plus having to use the restroom and all of these other things, are going to drive us out of our bodies as soon as we have the opportunity to leave them. Lots of people would leap at the chance to get out of their bodies if they could go plug into a virtual world.

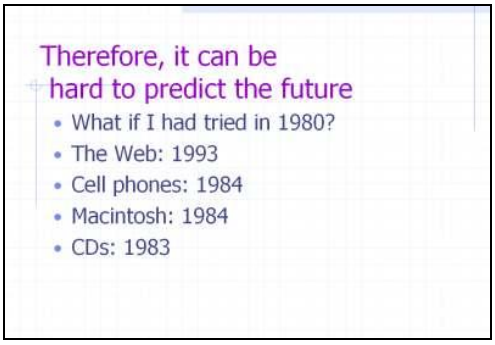

Image 3: Predicting the Future

We are all inhabiting human bodies right now. We all, though we don't realize it or consciously think about it every moment of every day, we all want out of our bodies. We would like to discard these vehicles that we currently use for transportation and we would like to replace them with something better.

What is going to happen? You can imagine me trying to get through airport security with a toy brain in a soda bottle. It's a fake brain. It's a toy and I actually brought the little packaging so I could show them it was just a toy brain.

This is what the next phase of technology will be, I think. I don't think we'll get to uploading fast enough. I think we will instead just discard our bodies, take our brain, and put them in containers that provide oxygen, nutrients, antibiotics and whatever else to keep us going. Then later we'll get to uploading.

The advantage to putting our brains in bottles is, first of all, we eliminate the whole trauma thing from falling down and getting in car wrecks and stuff. Next, there will be far less disease exposure because, it can be kept in a sterile facility and our bodies open us up to lots of diseases that our brains don't necessarily have to participate in.

The problem is that we will connect to virtual environments by living in bottles and have a lot more fun but the neurons still die. The estimate is about 30 million of your neurons die every year as you go through life.

We are going to want to store our brain in a permanent medium. That is where this whole idea of mind uploading [1] comes from. Back up our consciousness and run it on another medium where we don't have 30 million neurons dying every year.

Here are the basic facts on the brain, it's a liter and a half, it consumes 20 watts, it has 100 billion neurons, 100 trillion synapses, it's got a lot of atoms, maybe ten to the thirtieth atoms and it uses this basic component called a neuron, for its technology.

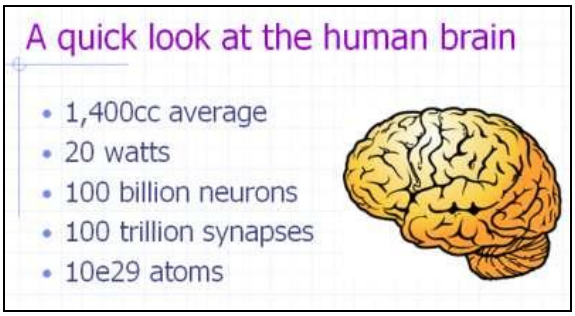

Image 6: Human Brain

The problem is how do you store this and then execute it in some other medium besides the current one? How do we take the patterns that are in it? The patterns are stored in at least three ways: the connections between the neurons, the formation of new synapses through experiences and then, microtubules.

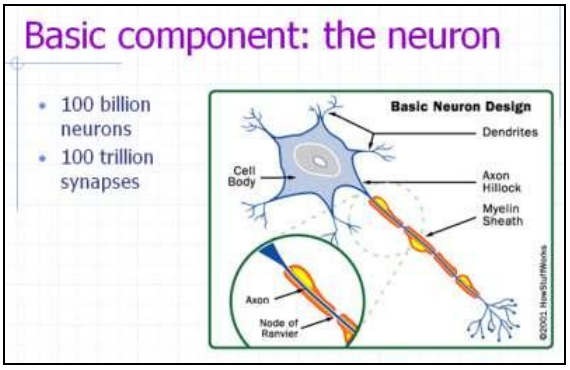

Image 7: The Neuron

Memory could be stored somehow but I don't think we even know how all memories are stored. We somehow take all that physiology, put it into some kind of computer medium, and then we have to figure out how to execute it, which may happen in two steps.

There are maybe ten to twenty different commonly discussed possible ways to do this. One is where they take the top of your skull off and they just probe your cortex with an electrode and you get really vivid memories of things that have happened in your life. Those memories play on your visual cortex and there are memories that play on your auditory cortex. So, one idea is to somehow probe the brain, basically scan through it memory by memory, and record images and sounds off the visual and auditory, and record it. That would be low fidelity but it would be a way of capturing the movies out of your brain.

The second way is neuron simulation of some sort. You have to somehow get inside the brain, probably destructively, and look at every single neuron and see how is it connected to all the other neurons around it, how are the synapses weighted, how are the synapses connected, and somehow tease that out of the structure.

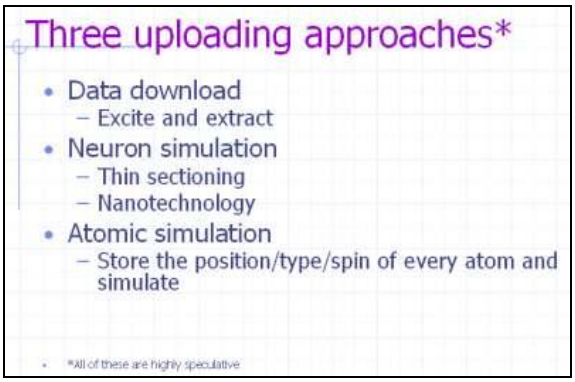

Image 8: Uploading Approaches

The two main ways of doing that have been proposed is either slicing the brain very gently and just scanning it in some way or injecting some kind of a nanotechnology entities into it that can look at and figure out how to emulate each neuron and either kill off that neuron and replace it, as that opportunity is available, or somehow stand alongside it and eventually have an image of every neuron in the whole brain that's being transmitted out by these nanobots.

Or, you go the whole distance and you somehow look at every single atom in this object and you store the type of atom, its location, it's bonding to neighboring atoms, someone mentioned (cork spin), and somehow take an atomic image of this.

Then the question is: how would you do that? No one has a really good idea right now but this Star Trek transporter room idea does offer one technology for doing that because it is already taking your entire body and turning it into an electromagnetic wave that can be transmitted to a planet surface.

That gives you an example of how speculative the technology is. No one has a good way of conceiving of how you would take something apart atom by atom and then simulating it into the atomic level and running it.

The thing that is so interesting is that all of these things are probably possible, perhaps within 40 years. In some form of this in some way within 40 years, that's extremely hard to imagine, yet probably true in the same way.

Going from the Wright Brothers' airplane to the B-52 was hard to imagine in 50 years.

I think back to that nice woman who was sitting on the airplane with me. In perhaps 40 years, that problem won't exist anymore. That is an amazing thing if it actually happens.

Endnotes

1. Mind Transfer/Mind Uploading - In transhumanism and science fiction, mind transfer (also referred to as mind uploading or mind downloading, depending on one's point of reference), whole body emulation, or electronic transcendence refers to the hypothetical transfer of a human mind to an artificial substrate. Wikipedia.org January 23, 2007 3:47PM EST

BIO

Marshall Brain, founder How Stuff Works

Marshall Brain is a well-known national speaker and consultant. Best known as the founder of How Stuff Works, he has also authored several books including, the *Robotic Nation* essays, the book *Manna*, and a book for teenagers entitled *The Teenager's Guide to the Real World*, now in its eighth printing and selected for the New York Public Library's prestigious Books for the Teen Age list. He is a member of the Academy of Outstanding Teachers at North Carolina State University, where he taught in the computer science department for 6 years.

The JOURNAL of GEOETHICAL NANOTECHNOLOGY

Volume 2, Issue 1
1st Quarter, 2007

The Role of AGI in Cybernetic Immortality

Ben Goertzel, Ph.D.

Ben, as Founder/CEO of both Biomind LLC and Novamente LLC, insightfully expounds upon the knowledge of and differences between Artificial Intelligence and Artificial General Intelligence toward human and cyber-immortality.

Varieties of Immortality

What do we mean by immortality? A number of different things are gathered into that word. I'm reminded of a famous quote from Woody Allen that some of you are probably familiar with: "I don't want to be immortal through my work; I want to be immortal by not dying."

People have referred to various types of immortality: Biological immortality, living forever in your body, which is the most straightforward type; cyber-immortality, immortality by perpetuating oneself in a computational medium different from the original. You could upload yourself into a robot; turn yourself into a program running on a space satellite, etc.

Then there are various forms of partial and limited immortality, which some of us get some limited satisfaction from – things like writing books or software programs, or producing children. These processes persist some of your patterns beyond your own lifetime, but you can question how much of your own awareness or identity is really perpetuated.

In fact there are some philosophical questions even with uploading: If you upload yourself, is it really you or is it just some guy who's trying to be you? If your upload steals your wife, how much satisfaction do you get from it?

Gradual uploading may provide a way to get around this problem --what if you upload 1 percent of your brain today, the next 1 percent the next day, the next 1 percent the next day, and so on. Then there's some continuity between your current embodiment and the next embodiment – there's more of a sense that it's the same self all along.

If you get rid of the original guy after you create your upload, then you don't have the "upload stole my wife" problem. But otherwise, if you don't kill your old meat embodiment, then even if you gradually upload, you still may be left with two of you, two separate streams of experience. And this is a whole new way of thinking, both personally, legally and societally. That you can have multiple guys who are all Ben Goertzels or all Martine Rothblatts or whoever.

I tend to be more aggressive than Bill Bainbridge[1], in that I don't think the uploads

are something you'll have to hide in outer space. But that's a whole political question that I won't get into. In this talk I'm going to delve further into the philosophical issues than the political ones.

Identity and Other Illusions

Thinking about the notions of identity and awareness in the context of uploading brings up the fairly well-known fact that these are, in a sense, illusory concepts. These are ways that we have evolved to fool ourselves about the nature of ourselves. Most of the time we think we're enacting free will; we're really enacting neural programs that our consciousness, the reflective part of our brain/mind, isn't aware of.

There's a whole history of work in cognitive neuroscience by Michael Gazzaniga[2] and a host of others, demonstrating that when people believe they're acting according to free will, often the decision was already made by some other part of their brain beforehand. I don't have time to go into that in detail but it's really fairly strong evidence.

And there's a good book by a guy named Thomas Metzinger called "Being No One,"[3] which integrates philosophical and neuropsychological evidence pointing to the conclusion that what we think of as our self, our identity, the phenomenal self, is a kind of neurologically constructed illusion.

It's a very useful illusion, thinking of ourselves as a coherent identity. Thinking of ourselves that way is useful. Thinking of ourselves as having a continuing stream of consciousness, of being fairly fully self-aware -- this is useful, but not necessarily accurate.

This leads to very interesting questions regarding uploading and the move from human to transhuman awareness. It may be that if you upload yourself and then improve yourself so that you have a better rational understanding of what is going on inside your own mind, this could lead to the loss of these illusions.

If will, awareness and self are, in most part, illusions that we construct because of our evolutionary heritage, and our limitations; then maybe, once we get smarter and more aware, we'll get rid of them. That gets back to Randal's earlier question of, you know, do we want subjective experience to be preserved?

Aspects of our subjective experience may come to seem quite idiotic to us, once we get a little smarter. And of course, being a good old American individualist, I would rather see each sentient mind able to make that choice for itself -- and if desired, to make multiple choices in parallel. Some minds may retain the illusion of being someone – the illusion of having will, and self, and self-consciousness -- and others may grow beyond this level.

I think these are all very interesting issues. I've explored some of these in a recent book called "The Hidden Pattern,"[4] which tries to present a patternist philosophy on mind. You can look at a mind as a system of patterns associated with some physical or computational system. A mind is a system of patterns that achieve goals by recognizing patterns in themselves and in the world -- and in that sense cyber-immortality is just a matter of the set of patterns that constitutes a given mind being replicated in some other means.

And if you look at mind as being about the emerging patterns rather than about the substrate, cyber-immortality is not really a big deal. On the other hand, you can also see that moving a mind to a different substrate which is more flexible, may allow the set of patterns that is the mind to evolve in a direction that it

could not have evolved in, in its original substrate.

Practicalities of Cyber-immortality

What about the practicalities of cyber-immortality. Well, one approach to cyber-immortality is a topic that we've already gone over, somewhat. You can scan the brain.

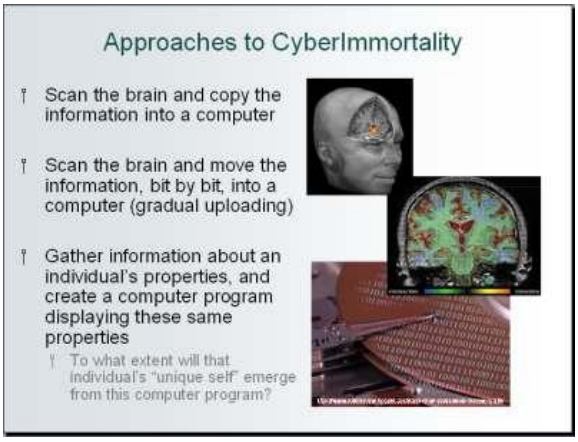

Image 1: Approaches to Cyberimmortality

You can copy the data into a computer. The copying process moves information bit by bit, giving some possibility for gradual uploading, providing some possibilities for a perceived continuity of consciousness.

Another possibility is to gather information about an individual's properties and create a computer program displaying these same properties. Take everything I ever wrote, everything I ever said that got recorded, movies of my behavior -- take all that data, put it all together in some program, based on a software program designed to feel like it's self-aware, designed to feel like it's Ben Goertzel.

I think this is an interesting notion. Maybe you could, with a sufficiently powerful AI, do this piecing together. In principle, maybe you could reconstitute Ben Goertzel from the traits, of everything that Ben did. But it's really, really hard. I'd rather not rely on it for my own immortality.

There's also the possibility, maybe, of a kind of quantum physics approach to reconstituting people. In principle, according to quantum physics, every macroscopic event is recorded in the universe itself. In the little perturbations of particles scattered through the cosmos. Quantum theory says that information is never actually destroyed – and you could compute the past from the present. In principle, with a sufficiently powerful computer, you could roll back time and figure out every single thing about every one of us. That would be a powerful way of doing uploading, but I'm not sure it would ever be feasible.

What are the obstacles between cyber-immortality and where we are now?

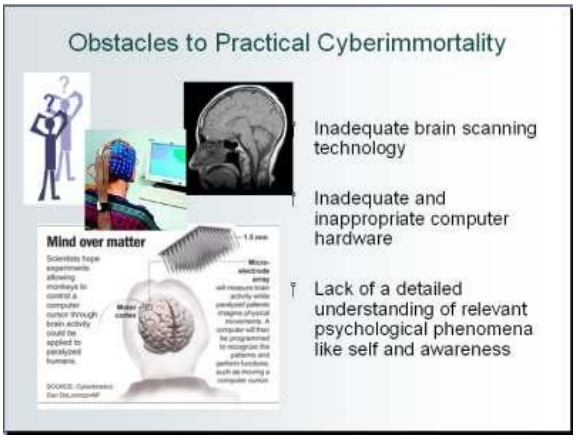

Image 2: Obstacles to Practical Cyberimmortality

Well -- basically everything. We don't have a scanned-in brain in enough detail. We don't have computer hardware that's good enough to receive a human-like intelligence; and we don't understand that much about self-awareness and other relevant phenomena to know if what we're doing will preserve what's important about ourselves. So right now it's a very valid and important goal to have -- as

something to guide our thinking and our research -- but we shouldn't delude ourselves that any of the component technologies are ready.

Using AI to Explore Cyber-immortality Issues

Now one thing that's occurred to me is that some of these issues, these philosophical and conceptual issues related to cyber-immortality and uploading, can actually be explored using artificial intelligences.

It's sort of funny to think about uploading an AI because an AI's already a digital thing. But, imagine you have an AI that you can talk to and that says it feels like it's self-aware. It says to you: "Hey Ben, I'm your AI. I'm conscious. I'm aware. How're you doing?"

Suppose you take that AI and then you copy it to a different medium -- say, a different kind of computer hardware. Then when it becomes smarter, what does that uploaded, improved AI say about the other one? Does it say, "Hey, you stole my hard drive"? Or does it feel like the same one? How many changes can you make to the AI's intelligence levels, to the AI's implementation -- and have it still feel like it is the same mind, the same identity.

It may well be possible to experiment with these ideas with AI programs with more flexibility than we can do with humans -- because with an AI, both we and it command a greater ability to inspect its own internals than exists with human beings. Potentially, we could discover disturbing things. We might discover that, in every case, if we double the intelligence of a system, it doesn't feel like its old self at all afterwards. It feels like it's a totally different thing. And if we did discover this, this would let us know that doubling our intelligence is basically equivalent to murdering our identity.

On the other hand, we might find that if intelligence is ramped up gradually, then there is a feeling of continuity in the emergent pattern constituting our phenomenal self – that is, our identity is preserved. And then we would know that we should upload ourselves more gradually, if we care about identity preservation.

I think it may wind up that we can explore a lot of these issues at the boundary of cognitive science, uploading, and philosophy with AI minds, rather than by experimenting with them initially on ourselves.

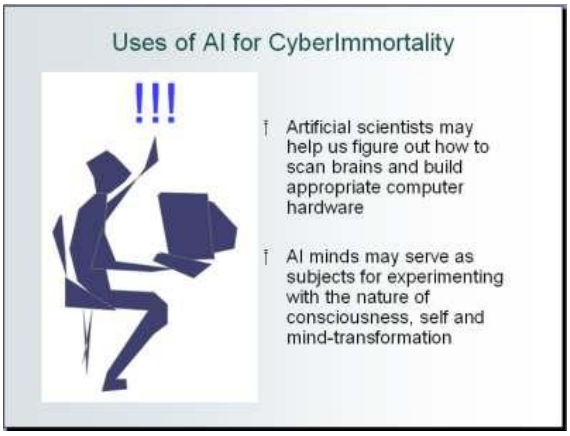

Image 3: Uses of AI for Cyberimmortality

Of course, whether this is the way things happen depends on whether AI progresses rapidly compared to the infrastructural technologies for human uploading.

All in all, I think there are two uses that AI technology may have to assist with cyber-immortality. One of them is that if AI proceeds rapidly, artificially intelligent scientists can help us figure out and solve these very difficult problems with uploading. They can help us figure how to scan brains, how to build better computer hardware. And the other issue is the one I just mentioned: AI minds may potentially serve as subjects for experiments with the nature of consciousness itself and mind transformation.

And there are ethical issues here. Once you've trained an AI mind, you may have a system that's a conscious intelligence just as much as we are. You know, it's not very nice just to mutate its consciousness, against its will. "I'm sorry; we are now going to make you mentally challenged. You didn't like it, huh? Too bad. You're just an AI. You have no rights." You don't want to do that.

On the other hand, it may well be the AI mind will willingly participate in appropriate experiments on its mind -- as I would personally, particularly, if as with an AI there would be opportunities to be rolled back to my prior state.

So I think AI can have an important role in cyber-immortality in these two different ways.

Approaches to AI

AI is an umbrella term – "artificial intelligence"used today to cover a lot of things. I don't think it's a terribly good term because, after all, an artifice is a tool and AI's may not want to be our tools. It may not be appropriate.

In the end if you view the physical universe as a kind of computing infrastructure, the distinction between artificial (i.e. computational) and biological intelligence comes to seem kind of arbitrary. But I'll accept the word AI because it's well known; people know what I mean when I use it.

There are various types of systems that can be grouped under this label of AI. One type is what I call a narrow AI system, which is I believe, a term I picked up from one of Ray Kurzweil's books.[5] What that refers to is that systems that are highly intelligent in some narrow domain. Deep Blue[6], the chess playing program is a good example of that. Or, the system created by Sebastian Thrun and his team at Stanford University recently which won the DARPA Grand Challenge for artificially intelligent automobile driving.[7] That's another example.

These are great programs; I'm very excited about them. They just do one particular thing. They don't have any reflective awareness. They don't have any understanding of context. But they do one thing very intelligently.

One of the big lessons in the history of AI research over the last 4 or 5 decades has been the small amount by which progress in narrow AI, actually contributes towards the goal of more general reflective AI. That wasn't really foreseen. I think in the 1960's it was thought that making programs like Deep Blue or Mathematica [8] or Google would be big steps toward getting generally intelligent programs that can really think. I don't think it quite works it out that way.

The narrow AI programs have provided useful tools and insights -- but it's not quite accurate to say they're stepping stones along the way to general AI. That's a valuable and interesting scientific lesson.

Another kind of AI system that's interesting to think about is a totally general intelligence. There's been some fascinating theoretical work done by some European computer scientists, Marcus Hutter [9] and Juergen Schmidhuber [10] and some of their PhD students in Switzerland. What they have proven, using some really complex mathematics, is essentially: If you have arbitrarily much computing resources, you can get an arbitrarily powerful general intelligence. That may seem obvious but to prove it rigorously took a lot of advanced mathematics.

This sort of theory is nice, but I don't think it's terribly useful in terms of making intelligent

systems that can do anything. Because the totally general AI's they describe require more computing power than exists in this physical universe (by the current estimates). I think their work is philosophically important, in terms of indicating that the real problem with AI is computational resources.

The whole challenge of AI is getting intelligent behavior given the limited processing time and memory that actually exist. If you have arbitrarily much computing resources, their theoretical work shows pretty nicely that you can get arbitrarily much intelligence, according to a pretty general and reasonable mathematical definition of what intelligence is supposed to be.

Finally, we have the sort of AI which interests me the most, which is what I call artificial general intelligence.

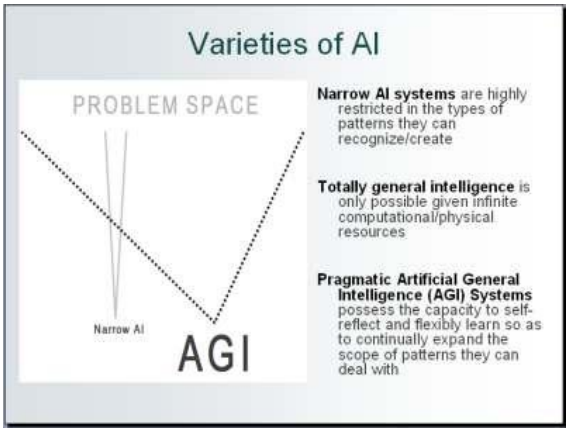

Image 4: Varieties of AI

What I mean is not that general intelligence can do anything, but intelligence that solves more than one problem – intelligence that is able to go into a context and figure out how to achieve its goals in that context autonomously and proactively.

As an example, rather than just playing one game like a narrow AI program like Deep Blue does, an AGI should be able to learn how to play a game by example and figure out the rules on its own. It should be able to do what a human can do: Go into a new country, learn the language, learn the customs and figure out how to represent information for itself. It's kind of a fuzzy definition when you start talking about general intelligence as opposed to narrow AI or totally general intelligence – its level of generality lies somewhere in between.

In practice, you can think about an AGI system as one that has the capacity to reflect on itself, to creatively learn and adapt. There's been distressingly little research in this area, in the AGI field. I think because it's more difficult, pragmatically than narrow AI research; and more difficult mathematically that totally general AI research. It's just hard. Yet of course, it's the most interesting thing in the long run.

I've been harping on the term "general intelligence," but there are also some other related terms. Actually it's been interesting to see that in the last few years there've been a number of workshops on the topic of the "human level intelligence," within mainstream AI conferences. I don't like the term "human level AI" very much because I think it sets the goal too low. I don't think humans are that intelligent in the scope of all possible minds. We should be setting our sights much higher than that. I also think "human level" is kind of ambiguous. What does it really mean? I understand what human-like means, but "human level" for a radically non-human intelligence is kind of poorly defined.

My colleague, Bruce Klein [11], this past May, helped me to organize a workshop on artificial general intelligence, which brought together various people from the futurist community together with a number of AI researchers from industry and academia. I think it was probably the first large scale collision between academic AI guys and radical futurists. Stan Franklin

[12] was there – he's fairly well known, from the University of Memphis; Sam Adams [13], leader of IBM's Joshua Blue project; and a number of other academic and industry AI researchers. They were fairly conservative guys with – and it was interesting to see them put together with but Eliezer Yudkowsky [14] and Hugo de Garis [15] and a bunch of the more outspoken visionaries in the AI futurist community. It was interesting how small the gap was between these various guys actually – many of the academic AI guys really had more of an interest in general intelligence and superhuman AI than they commonly liked to admit within the academic context.

There's also an edited volume which I'm co-editor of which gathers together a number of papers on general intelligence. It's called Artificial General Intelligence, published by Springer.

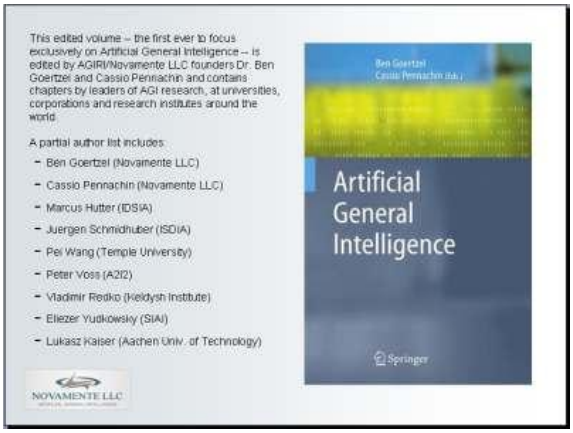

Image 5: Artificial General Intelligence Report

In 2007 IOS Press will publish the Proceedings of the AGI workshop.

Nick Cassimatis edited an issue of AI Magazine on the topic of human level intelligence recently. The field is building up a little bit of momentum. I have a feeling that somewhere within the next 5 to 15 years, to be conservative, you're going to see a renaissance of AGI, or human level intelligence research within the AI community.

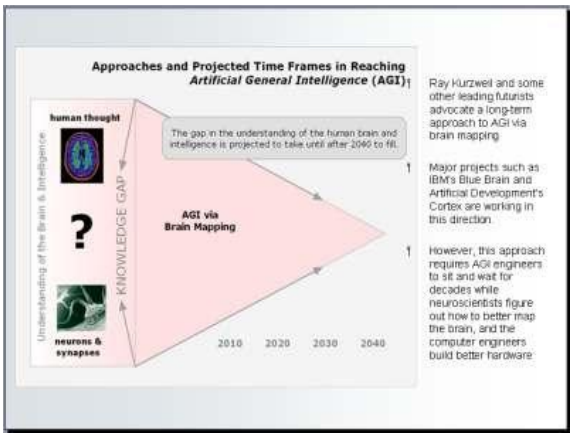

Image 6 -- [Click above for larger image]

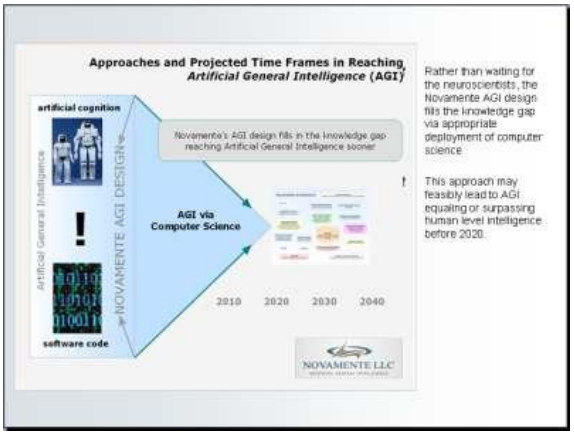

Image 7 -- [Click above for larger image]

I can see it building up because 5 years ago there were no workshops, special sessions, or anything on this kind of topic within mainstream AI conferences. Now at least there's some little corner being carved out for AGI. Out of 2,000 people at an AI conference, now there are at least 50 people who are interested in talking about human level general intelligence, which is progress.

All it's going to take is one exciting announcement – one announcement that someone has achieved some reasonably moderate level breakthrough in AGI. Then people will jump all over it, and the field will really explode.

Brain Modeling versus Computer Science Approaches to AGI

Next, I'm going to take 2 seconds to take a potshot at some of Ray Kurzweil's predictions regarding AI -- and then I'll spend a couple minutes just describing my own AI work at a very high level.

I think, in general, it's a lot easier to predict what will happen than predict when it will happen. This is very true, for example, in terms of software project estimation. Microsoft can't even tell how long it will take to make the next version of Windows. And it's also true with more long-term future logical prognostication.

There is the big question, for those that are interested in general intelligence, of what the route will be. Is it first going to be achieved through emulating the human brain? Not necessarily imitating it in detail but learning how the brain works and emulating those processes. Or is it going to be achieved through a more computer science approach? Where you take what's known about cognitive science as an inspiration and then use computer science algorithms and architectures to realize intelligence, rather than emulating what the brain does in detail.

My own view is that we just don't know enough about the brain. The most interesting and important parts of brain function -- higher level cognition -- are not understood at all. My guess is it's going to be a while before we know enough about the brain to use neuroscience to guide the creation of a general intelligence. My own prognostication, for what it's worth, is that this route would take until around 2040. I myself will be a disturbingly old man (and maybe even a disturbing old man!) before we get an AGI by emulating the human brain – unless some other route gets us to AGI first and helps us scan and map the brain and make better hardware.

My own feeling is that if a concerted effort is made in funding this area -- it's really not the case right now -- then artificial general intelligence via computer science methods can be achieved long before that. I'll take a number -- not quite out of a hat, but it will seem that way of context of this talk, and I don't have time to go into my reasoning – and propose 2020 as the date. Twenty, twenty has a nice ring to it, doesn't it? I'm taking that as a date that powerful AGI could quite possibly be achieved through computer science methods if we don't have to wait for the neuroscientists. And it could happen much sooner than that, I think. It could happen in five years from now, with my own Novamente project, if we got enough funding and everything went right.

A computer science based approach is a higher risk approach, in a sense. If you carry out predictive reasoning in a kind of plodding, methodological, conservative way, it's really obvious that if you map out what the brain does and emulate it in a machine, you've got to be able to make an AI that way. I mean there's the objection that maybe the brain uses macroscopic quantum effects in some weird way. But even then you just need to build a quantum computer instead of a classical computer.

The idea of making AGI by computer science is more risky. Maybe we're not smart enough, maybe the designs we think of will fail. On the other hand, although it's more risky, it also has more potential to proceed really fast because you don't have the huge overhead of waiting for the neuroscientists to map the brain. This is the approach that interests me most.

The Novamente Approach to AGI

My own approach to general intelligence in the last few years has centered around a software system called Novamante.[16] Novamente means new mind. It is also a Portuguese word for "again." And I chose the Portuguese term since a number of my software collaborators are actually based in Brazil rather than the U.S.

Novamente is a C++ software system which has been designed in a lot of detail. It's a big system. We're slowly plodding through the process of implementing and testing the various components. Maybe 40 percent complete in implementing the thing. And this has been a kind of spare-time background project since 2001. Just recently this year we now have three people full time dedicated to the project. We're staring to see a decent pace, although nowhere near the pace we would like to see.

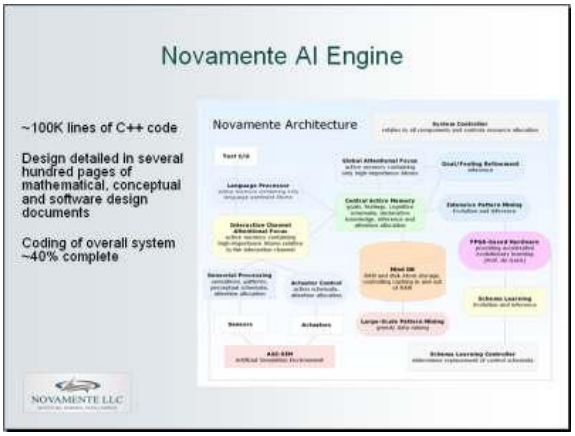

Image 8 -- [Click above for larger image]

Now some components of the system have been commercially deployed, in some software consulting projects that we've done in the areas of biology and natural language processing. But the process of using components of the system in these narrow AI consulting projects has been instructive regarding the big difference between AI and General AI. The bits and pieces of software we've used to help NIH (National Institute of health), INSCOM (US Armey Intelligence and Security Command) and other customers just really don't get to the essence. What their AI projects need is fairly simple stuff for pattern mining or language analysis and -- none of these customers so far has wanted to fund the long and difficult process of constructing a system that can reflect on itself and understand itself.

What we are doing to move toward general intelligence right now is to embody our AI systems in a 3D simulation world. The simulation world itself is based on an open source video game engine called Crystal Space. The AI controls a humanoid agent in the sim world, and the human teacher teaching the AI controls another humanoid agent. The idea is that you interact with the AI in this world and try to teach it stuff. You can chat with the AI in a little chat window; and it can walk around and pick things up and so forth.

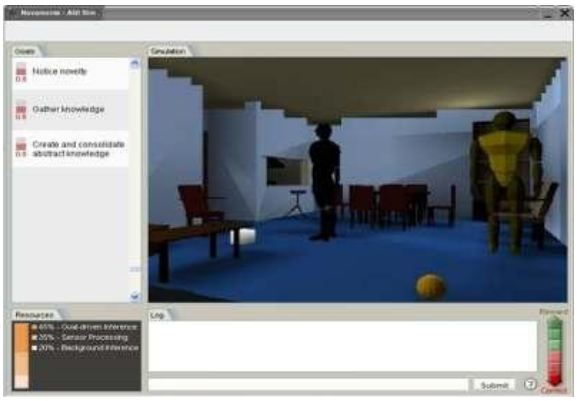

Image 9: Novamente AI Sim

This simulation world project is just getting started. It's still a bit buggy. The agent walks a bit awkwardly -- but it's OK, since it is a robot. And we haven't done much with language learning yet, but we're working it. Right now we're dealing mostly with very, very simple stuff like playing fetch; hiding an object and seeing if the AI remembers that it still exists (what Piaget [17] called object permanence).

The learning methodology is to try to build an artificial baby which learns everything it needs to know just based on its own experience and its interaction with you, and progressively gets smarter and smarter. We chose a simulation world rather than a physical world, largely for pragmatic reasons. It's lower cost. It's easier for a distributed team to deal with.

Ultimately, it would be nice to embody Novamente in a physical robot and make the simulation world a rather detailed simulation of the physical robot itself. Getting into an uploading vein, you could also make one of these simulated guys be a virtual Martine or a virtual Ben.

In the long run, you could use a massive simulated world like Second Life [18] as a vehicle here. You could have partial human uploads, baby AI's, and human-controlled avatars all interacting with each other -- and I think that's a great way for AI's to learn.

What makes Novamente different from other approaches to AGI? The real answer to that gets deep into the details. But on a more philosophical level, one thing I think makes the Novamente approach unique is that we pay more attention to the emergent structures of intelligence -- these things like self, free will, reflective awareness, and so on. I spent a lot of time trying to understand on a theoretical basis how these things can emerge from the lower level learning and knowledge representation infrastructure of Novamente. And I feel like most people have not taken that kind of approach.

There are AI systems that are based on logic -- logical reasoning – and don't pay much attention to self-organization, and emergence and complexity. There are approaches based on neural and evolutionary learning, which are great but don't deal with language and abstract reasoning hardly all. And there doesn't seem to be much understanding of how to make abstract reasoning emerge from these low-level structures.

Then there are integrated systems that integrate various modules together -- but in a kind of a plug and play way that doesn't give much thought to how they interoperate to produce emergent structured intelligence. I think it's really necessary to think about these high level structures: identity, self-awareness, long-term memory and how does it self organize? How to get these high-level properties of mind to emerge out of computer science and infrastructure is not obvious; and has been mostly what I've thought about over the last couple decades.

Getting back to the simulation world and teaching baby AI's: In terms of learning, I think a lot of Piaget's general framework, although obviously many of the details of his thinking need to be updated in accordance with the recent understanding of developmental psychology. I think about Novamente's progress in terms of developmental stages much like Piaget's. You can talk about an infantile stage, and then a concrete operational stage where we have a richer variety of mental representations and operations; a formal stage, where you can do abstract reasoning and hypothesis; and finally, the reflective stages and full-self understanding.

Piaget mostly talks about the first three stages. Later psychologists talk about post-formal thinking, involving deep reflection on the foundations of self and thought. AI can go even further than that; it can completely modify its own mind.

We are still at the infantile stage with Novamante. I think it is important to ascend that ladder methodically and to be sure the system has really mastered each stage before you go any further.

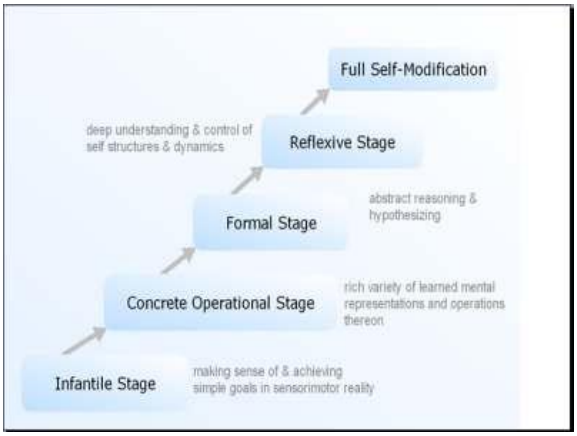

Image 9: Novamente Stages

The details of the Novamente architecture and how it represents knowledge shows the different parts of the system: short-term, long term memory, reasoning, learning, perception, and so on. It's a fairly complex system using cutting edge computer science, logic, evolutionary learning, and probability theory; using these things in a way to give rise to the emergent structures, of self, identity, and so forth.

The one point in terms of the AI architecture that I do want to harp on is the process I call map formation. What that means is as follows. Suppose you have a number of things in the system's memory -- say the nodes in the system's memory representing a bunch of related things like cat, mouse, tail, furry. Basically cat related stuff. Then there is a cognitive process in the Novamente system called map encapsulation, by which the system recognizes, "Hey, all these guys are used together. So let's make it a single concept to group all them together." Then a "cat-related" node would be created by the map encapsulation process, and could then enter into further reasoning.

This kind of process is a system recognizing patterns in what it does, and then embodying these patterns concretely and storing them explicitly within its own memory. It relates to the idea I raised in the comments to someone else's talk earlier, about a system taking its own implicit goals (what it is acting like its doing), and embodying those explicitly as in explicit goals (what it thinks it's doing). It also relates to the self. We recognize patterns of what we actually are, sometimes accurately, sometimes erroneously. We embody them as an explicit model of what we are. This process of recognizing patterns in yourself and embodying them explicitly and symbolically, gives rise to new patterns; this feedback is an important thing. I think if you get that feedback to work in an AI system you can get reflective awareness to work.

I think we're about one year away – if we get a bit more funding – from the creation of a fully functional, artificial infant.

Image 10: Goal For Year One After Project Funding

Once we get that, we can work toward ascending the next step of the ladder. I think the proper goal is to make an artificial baby in the simulation world, and I estimate we are about seven to ten man-years of programming and testing away from that. And then on to the next level in the Piagetan ladder.

I'm not saying that it's a trivial thing, it's hard. I'm just saying it's a palpable thing – it's a series of concrete, well charted set of steps. I

think this is a little different than AI through brain emulation, which is relying on a whole bunch of unknown stuff about mapping the brain, mind uploading, nanotechnology, etc.

If this kind of research program, either by me or others, is successful it will give us a lot of things. It will give us some amazing technologies, and a path to superhuman AGI. And it will also, like I mentioned before, give us a way to experiment with notions of identity, immortality, and self-modification.

A quick thanks to Bruce Klein, who helps me run Novamente, and to the excellent scientists and engineers who work with me on Novamente, helping me to try to bring the baby minds to life.

Endnotes

1. William Simms Bainbridge - (October 12, 1940 - present) is an innovative American sociologist who currently resides in Virginia. He is co-director of Human-Centered Computing at the National Science Foundation (NSF) and also teaches sociology as a part-time professor at George Mason University. He is also the first Senior Fellow to be appointed by the Institute for Ethics and Emerging Technologies. Bainbridge is most well known for his controversial work on the sociology of religion, however recently he has published work studying the sociology of video gaming. Wikipedia.org February 8, 2007 1:21 pm EST

2. Michael Gazzaniga - a professor of psychology at the University of California, Santa Barbara, where he heads the new SAGE Center for the Study of the Mind. In 1961, Gazzaniga graduated from Dartmouth College. In 1964, he received a Ph.D. in psychobiology from the California Institute of Technology, where he worked under the guidance of Roger Sperry, with primary responsibility for initiating human split-brain research. In his subsequent work he has made important advances in our understanding of functional lateralization in the brain and how the cerebral hemispheres communicate with one another. Wikipedia.org February 8, 2007 1:24 pm EST

3. Thomas Metzinger's "Being No One" http://www.philosophie.uni-mainz.de/metzinger/publikationen/BNO.pdf *February 8, 2007 1:16 pm EST*

4. Goertzel, Ben. The Hidden Pattern . BrownWalker Press, 2006. A Patternist Philosophy of Mind, "The Hidden Pattern presents a novel philosophy of mind, intended to form a coherent conceptual framework within which it is possible to understand the diverse aspects of mind and intelligence in a unified way." www.brownwalker.com *February 8, 2007 1:36 pm EST*

5. Ray Kurzweil - pronounced: [kɚz-waɪl]) (born February 12, 1948) is a pioneer in the fields of optical character recognition (OCR), text-to-speech synthesis, speech recognition technology, and electronic keyboard instruments. He is the author of several books on health, artificial intelligence, transhumanism, technological singularity, and futurism. Wikipedia.org February 8, 2007 1:54 pm EST

6. IBM Deep Blue - The computer system dubbed "Deep Blue" was the first machine to win a chess game against a reigning world champion (Garry Kasparov) under regular time controls. This first win occurred on February 10, 1996. Deep Blue - Kasparov, 1996, Game 1 is a famous chess game. However, Kasparov won 3 games and drew 2 of the following games, beating Deep Blue by a score of 4–2. The match concluded on February 17, 1996. Wikipedia.org February 8, 2007 1:56 pm EST

7. Sebastian Thrun & DARPA Grand Challenge – Director of Stanford University's Artificial

Intelligence Lab http://robots.stanford.edu/ *February 8, 2007 2:00 pm EST*

8. Mathematica - From simple calculator operations to large-scale programming and interactive-document preparation, Mathematica is the tool of choice at the frontiers of scientific research, in engineering analysis and modeling, in technical education from high school to graduate school, and wherever quantitative methods are used. http://www.wolfram.com/products/mathematica/index.html *Fenruary 8, 2007 2:03 pm EST*

9. Marcus Hutter, Ph.D. – an Associate Professor in the RSISE at the Australian National University in Canberra, Australia, and NICTA adjunct. His current interests are centered on reinforcement learning, algorithmic information theory and statistics, universal induction schemes, adaptive control theory, and related areas. http://www.hutter1.net/ *February 8, 2007 2:07 pm EST*

10. Juergen Schmidhuber - Prof. Jürgen Schmidhuber's main scientific ambition has been to build an optimal scientist, then retire. In 2028 they will force him to retire anyway. By then he shall be able to buy hardware providing more raw computing power than his brain. Will the proper self-improving software lag far behind? If so he'd be surprised. This optimism is driving his research on mathematically sound, general purpose universal learning machines and Artificial Intelligence, in particular, the New AI which is relevant not only for robotics but also for physics and music. http://www.idsia.ch/~juergen/ *February 8, 2007 2:19 pm EST*

11. Bruce Klein - born April 11, 1974, is the President of Novamente LLC, a privately held AI software company focused on Artificial General Intelligence (AGI). He also directs the non-profit Artificial General Intelligence Research Institute (AGIRI) and helped organize the first AGI Workshop May 2006. *http://www.answers.com/topic/bruce-klein February 8, 2007 2:25 pm EST*

12. Stan Franklin – Professor of Computer Science at the University of Memphis, TN http://www.msci.memphis.edu/~franklin *February 8, 2007 2:25 pm EST*

13. Sam S. Adams - an IBM Distinguished Engineer within IBM's Research Division, Watson research Center (Hawthorne), and leader of IBM's Joshua Blue Project, which applies ideas from complexity theory and evolutionary computational design to the simulation of mind on a computer. *domino.research.ibm.com February 8, 2007 2:41 pm EST and* *http://www.csupomona.edu/~nalvarado/PDFs/AAAI.pdf February 8, 2007 2:43 pm EST*

14. Eliezer S. Yudkowsky - an American self-proclaimed artificial intelligence researcher concerned with the Singularity, and an advocate of Friendly Artificial Intelligence. *Wikipedia.org February 8, 2007 2:46 pm EST*

15. Hugo de Garis - (born 1947, Sydney, Australia) became an associate professor of computer science at Utah State University. He is one of the more notable researchers in the sub-field of artificial intelligence known as evolvable hardware which involves evolving neural net circuits directly in hardware to build artificial brains. Wikipedia.org February 8, 2007 2:48 pm EST

16. Novamente - Novamente is a software product and development firm aimed at bridging the gap between narrow and general purpose Artificial Intelligence. Ongoing research brings the company's Novamente Cognition Engine closer each month to

powerful Artificial General Intelligence. Novamente.net February 8, 2007 3:32 pm EST

17. Jean Piaget - (August 9, 1896 – September 16, 1980) was a Swiss philosopher, natural scientist and developmental psychologist, well known for his work studying children and his theory of cognitive development. Wikipedia.org February 8, 2007 2:59 pm EST

18. Second Life - a 3-D virtual world entirely built and owned by its residents. Since opening to the public in 2003, it has grown explosively and today is inhabited by a total of 3,401,972 people from around the globe. Secondlife.com February 8, 2007 3:02 pm EST

BIO

Ben Goertzel, Ph.D,
CEO/CSO Novamente LLC

Involved in AI research and application since the late 80's. Former CTO of 120+ employee, thinking machine company, Webmind. PhD in mathematics from Temple University. Held several university positions in mathematics, computer science, and psychology, in the US, New Zealand and Australia. Author of 70+ research papers, journalistic articles and five scholarly books dealing with topics in cognitive sciences and futurism. Principle architect of the Novamente Cognition Engine.

The JOURNAL of GEOETHICAL NANOTECHNOLOGY

Volume 2, Issue 1
1st Quarter, 2007

The Ethics of Imagination: The Space Between Your Ears

Wrye Sententia, Ph.D.

Introduction by Dr. Martine Rothblatt

I had never really thought that there was an organization standing up for freedom of thought, and indeed before the neurosociety, why would one need an organization standing up for the freedom of thought.

Freedom of speech, it would be protection enough, because nobody could get into your thoughts other than by suppressing your speech or your behavior. And there are great organizations to look after freedom of speech and behavior. But as Zack has shown us, and ray before he, and others at this workshop.

We are at the cusp of a complete borderless meshing of all of our minds, and the ability to reach one's thoughts and for one's thoughts to reach others thoughts without ever slowing down to the speed of text is upon us.

Somebody needs to look after our freedom of thought, the Center for Cognitive Liberty & Ethics[1] that Wrye and her partner, Richard, have founded have been at the forefront of this effort.

The Ethics of Imagination: The Space Between Your Ears

This article will look at the concept of imagination and how imagination is key not only to the furtherance of many of the technologies that we see on a visionary horizon but also to fostering human consciousness in ethically meaningful ways, in ways that are sustainable as we move forward into the bumpy ride of the future.

Why do we need an ethics of imagination? Because ethics without imagination is dogma, and imagination without ethics is dangerous. In order to foster human consciousness, we must not only have an intention, but we must also have a capacity to imagine by improving the stalk of understanding, compassion, and indeed, empathy that goes with a socially conscious imagination.

Because I have found, in my personal experience, that a person who has an enhanced ability to empathize, that is to creatively imagine another persons circumstances is a person who engages in more ethical acts, in more conscientious actions and practices regardless of discipline or politics, whatever they may be. That is my plea.

The question is, how can we foster an ethical imagination for a wide spectrum of people, and anticipate ways to enhance the simulations even, of ethical behavior for Artificial Intelligence as we move into a long, extended future?

How can we do this without knowing in advance what sorts of changes we face in terms of human evolution, the massive shift in capabilities that we may see, and also in terms of societal evolution?

The way that I want to focus on is a turn inward, thinking about an emotional enhancement, one aspect of that which correlates to a more ethical thinking, more cognition that is grounded in empathy.

What is empathy? Well, the OED (oxford English Dictionary) tells us that the empathy is the power of projecting ones personality into the object of contemplation. If you look in the psychiatric literature, it is the capacity to understand what another person is experiencing from within the other's frame of reference.

Now this is key, because, if you think about sympathy, sympathy is a term that's existed since the 16th century, and it came out of a religious tradition of seeing one's human plight as common to other people.

In other words, it proposed a likeness between sympathizer and sympathized. So, the person who felt sympathy saw that you too were one of God's creations and in need of salvation. So there is implicit moral, religious overlay on sympathy.

Empathy, however, is only about a hundred years old. It is a word that came into use about a hundred years ago. Empathy presupposes difference. Its emphasis does not rely on feeling how the other person is like you, but really extrapolating. Using the virtual projection of the imagination to get to where someone else is at. Empathy builds on difference, sympathy builds on sameness.

What is ethics? Currently the idea of neuronanotechnology is very different for most people. And therefore coming to an ethical consensus on what ethics in relation to neuronanotechnology might be is not a foreseeable thing. Yet, I don't think we need to look for consensus in order to look for a more ethical process of analyzing new technologies in general, and neurotechnologies in particular.

Most of what people know today about nanotechnology is based on the confabulations of a popular imagination; things in the popular press; extravagant movies; things of this nature; doomsday scenarios; AI intelligence overtaking humanity; and then decimating any sort of consciousness that resembles a human entity.

These are fairly dystopic scenarios. However, what I argue is that rather than reject or distance ourselves from such negative or dystopic portrayals of a popular imagination of nanoscience or neuronanoscience, we should foster and encourage an interpretation of these cultural artifacts that actually increases the possibility for an empathic imagination, understanding difference through these creative venues. Society needs more tolerance, not less. More tolerance can be grown by encouraging this aspect of creative thinking.

If we can enhance what I'm calling an empathic imagination, we'll be able to enhance the ethical application of neuronanotechnology rather than relying on moral dictates or culturally and specific norms because you're not looking for a similarity, you're able to extrapolate to difference.

Virgil Ulam

Some of you may be familiar with what happened to Virgil Ulam. He was a genetics researcher in the 1980s in California, and he was working about the time that Eric Drexler's, *Engines of Creation,* came out.[2]

Ulam was fired from his company because on the side, outside of his legitimate company-sponsored research, he was experimenting with engineering cells. Jjust before he was fired, rather than loose his job, he decided to inject one of his last samples into his body in order to save the work.

Image 1

Now this may seem like a stupid thing to do and certainly Ulam's experience witnesses that effect, however, I think we can learn, again, something from his experience which points to the value of an empathic imagination.

Of course, Ulam expected to extract these cells from his body later, after he had left the secured company lab, but as it turns out, he wasn't able to, and the cells began to replicate. Except, rather that getting sick, Ulam actually found that his physical and mental properties -- his experience was improving, he was undergoing unexpected health benefits.

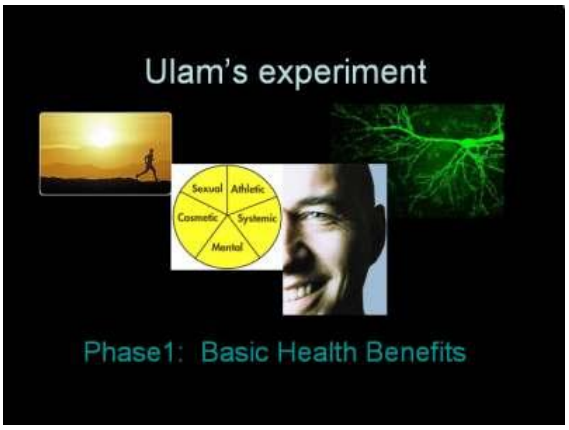

Image 2: Phase 1

I called this phase one, he felt a better agility, increased processing power, and improved mood and outlook, as well as improved memory recall, and other intellectual and physical capabilities.

A few weeks later, after he realized he couldn't extract the cells from his body, he began to report that he felt benefits well beyond his abilities and functions that might be considered normal. The engineered cells began to initiate life enhancing changes from correcting his twisted spinal column, to actually even improving his vision and his mental capabilities.

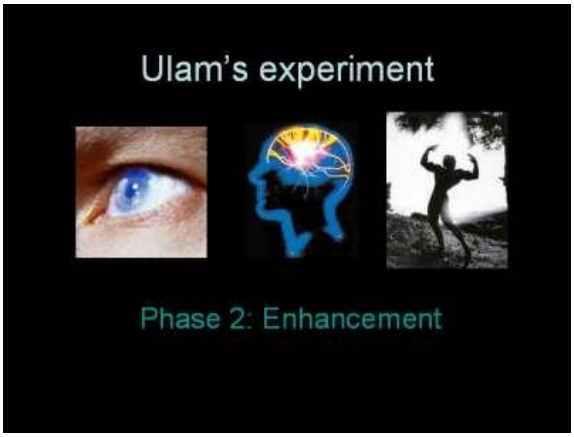

Image 3: Phase 2

But, after a few more weeks, Ulam documented shifts in his metabolism. He was becoming irritable and he was starting to

undergo negative consequences from his experiment.

Not long thereafter, Ulam found that the engineered cells, which had been previously kept out of his brain because of the blood brain barrier, had crossed into Ulam's brain where they began circulating and communicating electronically and synaptically with Ulam's neurons.

At this point, the cells began to convince Ulam of their superior world view. They did this through a series of different things: polite behavior, gentle reasoning, plus a dash of highly disruptive synaptic electrochemical behavior.

From this point on, things started to go badly for Ulam as a human. Now you are either saying, "Who the hell is Ulam? Wasn't he a mathematician? Ulam's crazy and so is Wrye," or you have recognized this for what it is, a science fiction plot.

"It is inner space, not outer, that needs to be explored."
J.G. Ballard (1962)

This story is from a book by Greg Bear called *Blood Music*, published in 1985, a year before Drexler's, *Engines of Creation*. *Blood Music* is a science fiction novel about engineered neuronanotechnology, or "smart cells," that eventually develop into an ever-expanding, conscious membrane.

Why am I sharing this with you? A few months ago, at the recent Singularity Summit at Stanford University, Chris Peterson, who's the Vice President of Public Policy at the Foresight Institute [3], said, "If you're trying to project the long-term future, and what you get sounds like science fiction, you might be wrong. But if it doesn't sound like science fiction, it's definitely wrong."

This calls attention to an unresolved conflict, and a complaint about discussions of nanoscale science and technology. Many critics complain that it is not so much science as science fiction that they're hearing in the place of science.

For instance, a Stanford University biophysicist, Steven Block, had criticized many nanoscientists, including Eric Drexler and the Foresight crowd, claiming that they have been influenced by, "laughable science fiction expectations."

Block complains that in order for real science to proceed, nanotechnologists ought to distance themselves from what he calls "the giggle factor." Certainly most university professors, industry researchers, government officials, have a strong insecurity to being taken for quacks.

Understandably, they try to distance themselves from such science fiction-esque scenarios. Or to put it more generously, they are concerned that by embracing some of the more visionary aspects of science, the more radical conjectures and hypotheses for nanotechnology, that they will encourage a hysteria or mania in the larger population.

Yet, I would say that it is just such speculative visions of future technology, in both its good and bad forms, in pursuit of innovative science or of a good story, that offer, through their ability to spark the imagination in positive ways, a way to catalyze a more comprehensive understanding of possibility and a more ethical future.

1984

For example -- this is what I call the "reality factor," -- Orwell's *1984* book. When that came out in 1949, George Orwell offered then, and it is still applicable today, a way for people to imagine a society that was laboring under

the totalitarian use of surveillance technologies.

And as a student in one of my UC Davis science fiction classes said recently, "Orwellian" has become its own adjective, and even if you never read the book, you know what it means when someone says that the government's NSA Surveillance Program is Orwellian.

Now you may be thinking, okay, but that's a very negative view that fifty years later we're still hearing, Orwell, Orwell, Orwell. But, it is just such a dystopic portrayal of the technology in a fictional book that allows the public today to rally around an outcry over the unethical use of a particular technology.

1984 allows the public a shorthand way to think about abuses, or government snooping and invasions of privacy. Even if they don't understand what cryptography might be, or how electronic data mining impacts their life in a daily way. It begins that shorthand imaginative use of a novel to impact a larger social society, or larger social conditions.

Another way to think about it is this: there's a person who writes for Scientific American fairly regularly. His name is Gary Stix. He is a vocal critic of nanotechnology and he has complained that Eric Drexler's writings are similar to the scientific romances of Jules Verne[4], or H.G. Wells [5].

And, that you can't find, "real" technology in speculative science. But when Stix says this, he's missing the point, because at the turn of the last century, Jules Verne and H.G. Wells were highly influential in stimulating an interest and the pursuit of innovative technologies and science.

The exploration that went with it, the positive search, was catalyzed by that. And, in the same way, in the 1980s, Drexler's *Engines of Creation* was highly influential in impacting not only the science, but also the science fiction of nanotechnology.

I'm making an appeal to embrace, rather than reject, the speculations in science, particularly these nano-fiction narratives that can inspire an ethics–related discourse of new technologies and applications.

Back to *Blood Music*: what happened to Ulam? We left Ulam with a smart cell circulating in his head; where they'd succeeded in convincing him of their lyrical harmony, of their blood music, their superior collective world view.

Now, ultimately the smart cells spread out from Ulam's body through his bath water and dominate, or take over other humans in a quest to convert--in the sense of convincing, but also in the sense of altering other humans.

At a cellular level, the cells take over the biological and social environments to which they are exposed, much like a virus. However, they radically restructure the human race in an evolutionary scenario, and in this scenario, humanity is corralled from its separate, autonomous beings, into an intelligent biomass.

It ends up becoming this sort of sheeny, phosphorescent, consciousness skin that spreads out over all of North America, covering the entire planet and eventually floating off into space as a conscious, thinking membrane.

Now, this is exactly the kind of scary scenario that Joachim Schumer has cited in his recent book that just came out a couple months ago. He's documented that it's just such grand; far-flung visions of nanotechnology that people mainly associate with the science, and which fascinates them, but also terrifies them.

Most science fiction commentators on *Blood Music* see this as a horror story of technology run amuck. For example, Dan Danillo writes: "Greg Bear's *Blood Music* takes the horror of exponentially, self replicating, intelligent nanomachines to its ultimate extreme, the termination of the natural world."[6]

However, I think it is just such a radically other vision, through the perspective and acceptance of such a vastly different form of technohuman existence in a fictional future that provides a safe and useful way for the public to entertain the possibility of future social and ethical implications of new technologies in a non-threatening way.

Such a story as *Blood Music* invites readers to reassess their own position or perspective; stretching not their skin, but their consciousness. With *Blood Music*, where it asks to consider what is the high price of such a transition of fully integrated, interactive, and a harmonious smart culture; loss of individuality, the loss of self, but also the loss of selfishness.

At the same time that there's this negative depiction, readers of *Blood Music* are also invited to entertain the idea of this ever-expanding cellular colony of the next, or even desirable step, an evolutionary step, for the human species.

I'm not staying that we need to shuck our humanity, and embrace a high mind in the form of a skin-like planet. But, with the freedom to imagine, we are invited and even compelled to relate to a different kind of consciousness, which I think can lead to a different and more comprehensive kind of ethics.

This takes us back to the empathy/sympathy issue. If you, as a reader, sympathize with humanity, then yes it is a horror story because you don't see the possibility for a resonant other.

However, if you empathize with the smart cells, and you're invited to do that too as a reader of *Blood Music*, then you can imagine the value of a harmonious culture; a global intelligence that's a viable alternative to overcoming some of those aspects of human culture that are found lacking.

From the vantage point of the scientist, or the nanotechnologist, rather than trying to dismiss some of the radically or potentially threatening science fiction visions, I invite the scientific community to engage these texts in ways that will benefit from such a radical perspective shifting.

Now, why do I think that this is a sustainable argument? Because, in the 18th century, it turns out, when the genre of the novel was just beginning, fictional narratives played a key role in the emergence of what was then a new idea: the then new political and legal concept of human rights.

Lynn Hunt is a professor of history at UCLA, and she's argued that the widespread reading of the new genre of novels in 1740s and 1750s was responsible for creating individual experiences and that an inward experience inspired empathy, and made possible these new social and political formations that the French Revolution solidified.

Lynn Hunt explains that rather than reading the dry political tracts of the time by the likes of Diderot [7] and Rousseau [8], people were widely reading these novels that encapsulated, or incarnated their radical, political ideas in fictional form.

It was through a fictional resonance with characters that people came to understand and appreciate that difference of class, did not

need to mean difference in rights. Specifically, Hunt explained that the people reading these identified with protagonists who were very often a poor servant girl, it was sort of the trope in the 18th century that had all these novels about poor servant girls being exploited, and sort of taken advantage of.

But upper-class men, military officers, the upper echelons of the 18th century found themselves strongly identifying with these female servant girl characters because the books were part of a new genre that was experimenting with that kind of empathic identification in characterization.

Even though they had little in common with the characters in these novels, it led in part to the acceptance of the belief, or the belief and then the acceptance, of universal human rights.

It is because science fiction scenarios create narratives rich in imagined possibilities, rich in the imagined consequences that they offer a unique way today for people to relate to new science and to understand some of the sociopolitical issues that could be attended with that.

I should also mention that in the 18th century, the novel, because it was this new narrative form, was considered lowbrow literature. Science fiction often gets classed as a popular pulp fiction-esque kind of literature. In the 18th century, the novel was operating in the same way.

Tomorrow's Ethics

One of the framing premises for my talk is that tomorrow's ethics and public policy can be exponentially enhanced by applying today's tools for greater empathy. Even as we anticipate other forms of techno-social catalysts in the area of ethics, we can look forward to fully immersive virtual realities.

I'd like to see fully immersive emotive realities and more finely tuned neuropharmaceuticals. There's a class of drugs today that's known as empathogens; in other words, awakening empathy within, generating empathy.

And there are certainly other un-dreamt of possibilities in terms of new neuro-nano applications that could foster imagination and even possibilities for other forms of conscious existence.

My point about a social evolution and an empathic society is an analog to raise Ray Kurzweil's model of technological accelerating returns. In explaining that, looking at biological evolution or technological evolution, you can see that today's rate of progress is often confused with these linear projections of the past and over-the-shoulder looks at how things were, so that's how things are going to be.

It is mistaken, and I see that social evolution is sometimes, particularly discussions of ethics, stymied by that same over-the-shoulder look. Rather than anticipating how things *could be*, and again--I'm not saying we should turn into a skin-like planet--but rather than anticipating ways to expand and enhance our empathy, people look to the past.

One area, , that we see this over-the-shoulder look, is the way that the law operates by precedent. And Richard Glen Boire, my partner at the Center for Cognitive Liberty & Ethics, echoing Marshall McLuhan [9], said that not only do we drive culture forward by looking through a rear-view mirror, but the law moves forward by looking through a rear-view mirror.

There is this enormous emphasis on precedent and tradition in our culture, and yet the law is

finding today -- the law as an entity -- the legal system is discovering that in an age of interactive and converging digital technologies, that looking to the past in order to figure out how to operate, and how society might be in the future, doesn't work well when you entertain radical technologies.

In order to exponentially enhance ethics, we need to also enhance the legal rights that go with them. Things like freedom of expression and freedom of thought will need innovative ideas and investments in speculative social, political, and technological possibilities -- not a rejection of them.

In terms of neuro-nanotechnology, we will need to envision protections that will ensure both a freedom to use the beneficial applications, as well as to protect the future Amish, a freedom from the coercive measures of potentially neuronanotechnology, and Zack touch on many of those.

I'll close with an appeal to cognitive liberty for the preservation of human consciousness. Cognitive liberty is concerned with fostering a right to think, particularly without governmental interference and in securing the right to explore, expand, and enhance your imagination with, or without, neurotechnologies.

Imagination, I feel, is a strong aspect that makes freedom of thought meaningful; hence the focus for today's talk. And, broadly, the Center for Cognitive Liberty and Ethics seeks to protect and foster a diversity of thinking. We can encourage a biodiversity, with a strong emphasis on our capacity to think. I will close with an Einstein quote.

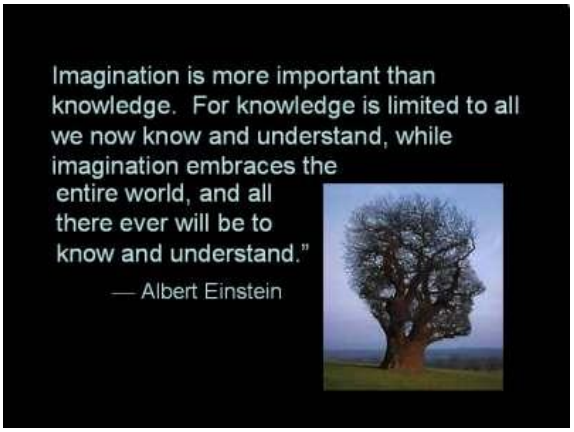

Image 3

Endnotes

1. Center for Cognitive Liberty & Ethics - a network of scholars elaborating the law, policy and ethics of freedom of thought. Our mission is to develop social policies that will preserve and enhance freedom of thought into the 21st century. Cognitiveliberty.org February 12, 2007 12:31 pm EST

2. Engines of Creation (The Coming Era of Nanotechnology) – (Marvin Minsky) "[A]n enormously original book about the consequences of new technologies. It is ambitious and imaginative and, best of all, the thinking is technically sound." Drexler, K. Eric. Engines of Creation. New York: Anchor Books, 1986.

3. Foresight Institute - "[A] leading think tank and public interest institute on nanotechnology. Founded in 1986, Foresight was the first organization to educate society about the benefits and risks of nanotechnology. At that time, nanotechnology was a little-known concept." Foresight.org February 12, 2007 12:56 pm EST

4. Jules Gabriel Verne - (February 8, 1828–March 24, 1905) was a French author and a pioneer of the science-fiction genre best known for novels such as Twenty Thousand Leagues

Under The Sea (1870), Journey To The Center Of The Earth (1864), and Around the World in Eighty Days (1873). Verne wrote about space, air, and underwater travel before air travel and submarines were invented, and before practical means of space travel had been devised. He is the third most translated author in the world, according to Index Translationum. Some of his books have been made into films. Verne, along with Hugo Gernsback and H. G. Wells, is often popularly referred to as the "Father of Science Fiction". Wikipedia.org February 12, 2007 1:00 pm EST

5. H.G. Wells - (September 21, 1866 – August 13, 1946), better known as H. G. Wells, was an English writer best known for such science fiction novels as The Time Machine, The War of the Worlds, The Invisible Man, and The Island of Doctor Moreau. He was a prolific writer of both fiction and non-fiction, and produced works in many different genres, including contemporary novels, history, and social commentary. He was also an outspoken socialist. His later works become increasingly political and didactic, and only his early science fiction novels are widely read today. Wells, along with Hugo Gernsback and Jules Verne, is sometimes referred to as "The Father of Science Fiction". Wikipedia.org February 12, 2007 1:02 pm EST

6. Denis Diderot - (October 5, 1713 – July 31, 1784) was a French philosopher and writer. He was a prominent figure in the Enlightenment, and was the editor-in-chief of the famous Encyclopédie. Wikipedia.org February 12, 2007 1:43 pm EST

7. Jean-Jacques Rousseau - 1712 – July 2, 1778) was a Genevan philosopher of the Enlightenment whose political ideas influenced the French Revolution, the development of socialist theory, and the growth of nationalism. Rousseau also made important contributions to music both as a theorist and as a composer. With his Confessions and other writings, he practically invented modern autobiography and encouraged a new focus on the building of subjectivity that would bear fruit in the work of thinkers as diverse as Hegel and Freud. His novel Julie, ou la nouvelle Héloïse was one of the best-selling fictional works of the eighteenth century and was important to the development of romanticism. Wikipedia.org February 12, 2007 1:44 pm EST

8. Daniel Dinello – author of Technophobia! Science Fiction Visions of Posthuman Technology books.google.com February 12, 2007 1:15 pm EST

9. Herbert Marshall McLuhan - CC (July 21, 1911 - December 31, 1980) was a Canadian educator, philosopher, and scholar-- a professor of English literature, a literary critic, and a communications theorist. McLuhan's work is viewed as one of the cornerstones of the study of media ecology. McLuhan is well-known for coining the expressions "the medium is the message" and the "global village". Perhaps the most celebrated English teacher of the twentieth century, McLuhan was a fixture in media discourse from the late 1960s to his death and he continues to be an influential and controversial figure. Years after his death he was named the "patron saint" of Wired magazine. Wikipedia.org February 12, 2007 1:57 pm EST

Wrye Sententia, Ph.D.

Wrye Sententia is director of the Center for Cognitive Liberty and Ethics (CCLE), a nonprofit research, policy, and public education center working to advance and protect freedom of thought into the 21st century. Dr. Sententia has guided the CCLE in sponsoring the National Science Foundation's initiatives aimed at "Converging Technologies for Improving Human Performance." In 2002, Sententia provided comments to the appointed President's Council on Bioethics in Washington D.C., on the topic of cognitive enhancement technologies and in October 2004 debated members of the Council on the democratic values of the US Declaration of Independence in relation to emergent enhancement biotechnologies and human freedom.