One Half a Manifesto

By Jaron Lanier, *Edge*, 11 November 2000

I'll here share my thoughts with the respondents of edge.org, many of whom are, as much as anyone, responsible for this revolution, one which champions the assent of cybernetic technology as culture.

The dogma I object to is composed of a set of interlocking beliefs and doesn't have a generally accepted overarching name as yet, though I sometimes call it "cybernetic totalism". It has the potential to transform human experience more powerfully than any prior ideology, religion, or political system ever has, partly because it can be so pleasing to the mind, at least initially, but mostly because it gets a free ride on the overwhelmingly powerful technologies that happen to be created by people who are, to a large degree, true believers.

Edge readers might be surprised by my use of the word "cybernetic". I find the word problematic, so I'd like to explain why I chose it. I searched for a term that united the diverse ideas I was exploring, and also connected current thinking and culture with earlier generations of thinkers who touched on similar topics. The original usage of "cybernetic", as by Norbert Weiner, was certainly not restricted to digital computers. It was originally meant to suggest a metaphor between marine navigation and a feedback device that governs a mechanical system, such as a thermostat. Weiner certainly recognized and humanely explored the extraordinary reach of this metaphor, one of the most powerful ever expressed.

I hope no one will think I'm equating Cybernetics and what I'm calling Cybernetic Totalism. The distance between recognizing a great metaphor and treating it as the only metaphor is the same as the distance between humble science and dogmatic religion.

Here is a partial roster of the component beliefs of cybernetic totalism:

1) That cybernetic patterns of information provide the ultimate and best way to understand reality.
2) That people are no more than cybernetic patterns.
3) That subjective experience either doesn't exist, or is unimportant because it is some sort of ambient or peripheral effect.
4) That what Darwin described in biology, or something like it, is in fact also the singular, superior description of all creativity and culture.
5) That qualitative as well as quantitative aspects of inform
    1. ation systems will be accelerated by Moore's Law.

And finally, the most dramatic:

6) That biology and physics will merge with computer science (becoming biotechnology and nanotechnology), resulting in life and the physical universe becoming mercurial; achieving the supposed nature of computer software. Furthermore, all of this will happen very soon! Since computers are improving so quickly, they will overwhelm all the other cybernetic processes, like

> people, and will fundamentally change the nature of what's going on in the familiar neighborhood of Earth at some moment when a new "criticality" is achieved- maybe in about the year 2020. To be a human after that moment will be either impossible or something very different than we now can know.

During the last twenty years a stream of books has gradually informed the larger public about the belief structure of the inner circle of Digerati, starting softly, for instance with Godel, Escher, Bach, and growing more harsh with recent entries such as The Age of Spiritual Machines by Ray Kurtzweil.

Recently, public attention has finally been drawn to #6, the astonishing belief in an eschatological cataclysm in our lifetimes, brought about when computers become the ultra-intelligent masters of physical matter and life. So far as I can tell, a large number of my friends and colleagues believe in some version of this immanent doom.

I am quite curious who, among the eminent thinkers who largely accept some version of the first five points, are also comfortable with the sixth idea, the eschatology. In general, I find that technologists, rather than natural scientists, have tended to be vocal about the possibility of a near-term criticality. I have no idea, however, what figures like Richard Dawkins or Daniel Dennett make of it. Somehow I can't imagine these elegant theorists speculating about whether nanorobots might take over the planet in twenty years. It seems beneath their dignity. And yet, the eschatologies of Kurtzweil, Moravec, and Drexler follow directly and, it would seem, inevitably, from an understanding of the world that has been most sharply articulated by none other than Dawkins and Dennett. Do Dawkins, Dennett, and others in their camp see some flaw in logic that insulates their thinking from the eschatological implications? The primary candidate for such a flaw as I see it is that cyber-armageddonists have confused ideal computers with real computers, which behave differently. My position on this point can be evaluated separately from my admittedly provocative positions on the first five points, and I hope it will be.

Why this is only "one half of a manifesto": I hope that readers will not think that I've sunk into some sort of glum rejection of digital technology. In fact, I'm more delighted than ever to be working in computer science and I find that it's rather easy to adopt a humanistic framework for designing digital tools. There is a lovely global flowering of computer culture already in place, arising for the most independently of the technological elites, which implicitly rejects the ideas I am attacking here. A full manifesto would attempt to describe and promote this positive culture.

I will now examine the five beliefs that must precede acceptance of the new eschatology, and then consider the eschatology itself.

Here we go:

Cybernetic Totalist Belief #1: That cybernetic patterns of information provide the ultimate and best way to understand reality.

There is an undeniable rush of excitement experienced by those who first are able to perceive a phenomenon cybernetically. For example, while I believe I can imagine

what a thrill it must have been to use early photographic equipment in the 19th century, I can't imagine that any outsider could comprehend the sensation of being around early computer graphics technology in the nineteen-seventies. For here was not merely a way to make and show images, but a metaframework that subsumed all possible images. Once you can understand something in a way that you can shove it into a computer, you have cracked its code, transcended any particularity it might have at a given time. It was as if we had become the Gods of vision and had effectively created all possible images, for they would merely be reshufflings of the bits in the computers we had before us, completely under our command.

The cybernetic impulse is initially driven by ego (though, as we shall see, in its end game, which has not yet arrived, it will become the enemy of ego). For instance, Cybernetic Totalists look at culture and see "memes", or autonomous mental tropes that compete for brain space in humans somewhat like viruses. In doing so they not only accomplish a triumph of "campus imperialism", placing themselves in an imagined position of superior understanding vs. the whole of the humanities, but they also avoid having to pay much attention to the particulars of culture in a given time and place. Once you have subsumed something into its cybernetic reduction, any particular reshuffling of its bits seems unimportant.

Belief #1 appeared on the stage almost immediately with the first computers. It was articulated by the first generation of computer scientists; Weiner, Shannon, Turing. It is so fundamental that it isn't even stated anymore within the inner circle. It is so well rooted that it is difficult for me to remove myself from my all-encompassing intellectual environment long enough to articulate an alternative to it.

An alternative might be this: A cybernetic model of a phenomenon can never be the sole favored model, because we can't even build computers that conform to such models. Real computers are completely different from the ideal computers of theory. They break for reasons that are not always analyzable, and they seem to intrinsically resist many of our endeavors to improve them, in large part due to legacy and lock-in, among other problems. We imagine "pure" cybernetic systems but we can only prove we know how to build fairly dysfunctional ones. We kid ourselves when we think we understand something, even a computer, merely because we can model or digitize it.

There is also an epistemological problem that bothers me, even though my colleagues by and large are willing to ignore it. I don't think you can measure the function or even the existence of a computer without a cultural context for it. I don't think Martians would necessarily be able to distinguish a Macintosh from a space heater.

The above disputes ultimately turn on a combination of technical arguments about information theory and philosophical positions that largely arise from taste and faith.

So I try to augment my positions with pragmatic considerations, and some of these will begin to appear in my thoughts on...

Belief #2: That people are no more than cybernetic patterns

Every cybernetic totalist fantasy relies on artificial intelligence. It might not immediately be apparent why such fantasies are essential to those who have them. If

computers are to become smart enough to design their own successors, initiating a process that will lead to God-like omniscience after a number of ever swifter passages from one generation of computers to the next, someone is going to have to write the software that gets the process going, and humans have given absolutely no evidence of being able to write such software. So the idea is that the computers will somehow become smart on their own and write their own software.

My primary objection to this way of thinking is pragmatic: It results in the creation of poor quality real world software in the present. Cybernetic Totalists live with their heads in the future and are willing to accept obvious flaws in present software in support of a fantasy world that might never appear.

The whole enterprise of Artificial Intelligence is based on an intellectual mistake, and continues to expensively turn out poorly designed software as it is re-marketed under a new name for every new generation of programmers. Lately it has been called "intelligent agents". Last time around it was called "expert systems".

Let's start at the beginning, when the idea first appeared. In Turing's famous thought experiment, a human judge is asked to determine which of two correspondents is human, and which is machine. If the judge cannot tell, Turing asserts that the computer should be treated as having essentially achieved the moral and intellectual status of personhood.

Turing's mistake was that he assumed that the only explanation for a successful computer entrant would be that the computer had become elevated in some way; by becoming smarter, more human. There is another, equally valid explanation of a winning computer, however, which is that the human had become less intelligent, less human-like.

An official Turing Test is held every year, and while the substantial cash prize has not been claimed by a program as yet, it will certainly be won sometime in the coming years. My view is that this event is distracting everyone from the real Turing Tests that are already being won. Real, though miniature, Turing Tests are happening all the time, every day, whenever a person puts up with stupid computer software.

For instance, in the United States, we organize our financial lives in order to look good to the pathetically simplistic computer programs that determine our credit ratings. We borrow money when we don't need to, for example, to feed the type of data to the programs that we know they are programmed to respond to favorably.

In doing this, we make ourselves stupid in order to make the computer software seem smart. In fact we continue to trust the credit rating software even though there has been an epidemic of personal bankruptcies during a time of very low unemployment and great prosperity.

We have caused the Turing test to be passed. There is no epistemological difference between artificial intelligence and the acceptance of badly designed computer software.

My argument can be taken as an attack against the belief in eventual computer sentience, but a more sophisticated reading would be that it argues for a pragmatic advantage to holding an anti-AI belief (because those who believe in AI are more likely

to put up with bad software). More importantly, I'm hoping the reader can see that Artificial Intelligence is better understood as a belief system instead of a technology.

The AI belief system is a direct explanation for a lot of bad software in the world, such as the annoying features in Microsoft Word and PowerPoint that guess at what the user really wanted to type. Almost every person I have asked has hated these features, and I have never met an engineer at Microsoft who could successfully turn the features completely off on my computer (running Mac Office '98), even though that is supposed to be possible.

Belief #3: That subjective experience either doesn't exist, or is unimportant because it is some sort of ambient or peripheral effect.

There is a new moral struggle taking shape over the question of when "souls" should be attributed to perceived patterns in the world.

Computers, genes, and the economy are some of the entities which appear to Cybernetic Totalists to populate reality today, along with human beings. It is certainly true that we are confronted with non-human and meta-human actors in our lives on a constant basis and these players sometimes appear to be more powerful than us.

So, the new moral question is: Do we make decisions solely on the basis of the needs and wants of "traditional" biological humans, or are any of these other players deserving of consideration?

I propose to make use of a simple image to consider the alternative points of view. This image is of an imaginary circle that each person draws around him/herself. We shall call this "the circle of empathy". On the inside of the circle are those things that are considered deserving of empathy, and the corresponding respect, rights, and practical treatment as approximate equals. On the outside of the circle are those things that are considered less important, less alive, less deserving of rights. (This image is only a tool for thought, and should certainly not be taken as my complete model for human psychology or moral dilemmas.) Roughly speaking, liberals hope to expand the circle, while conservatives wish to contract it.

Should computers, perhaps at some point in the future, be placed inside the "circle of empathy"? The idea that they should is held close to the heart by the Cybernetic Totalists, who populate the elite technological academies and the businesses of the "new economy".

There has often been a tender, but unintended humor in the argumentative writing by advocates of eventual computer sentience. The quest to rationally prove the possibility of sentience in a computer (or perhaps in the internet), is the modern version of proving God's existence. As is the case with the history of God, a great many great minds have spent excesses of energy on this quest, and eventually a cybernetically-minded 21st century version of Kant will appear in order to present a tedious "proof" that such adventures are futile. I simply don't have the patience to be that person.

As it happens, in the last five years or so arguments about computer sentience have started to subside. The idea is assumed to be true by most of my colleagues; for them, the argument is over. It is not over for me.

I must report that back when the arguments were still white hot, it was the oddest feeling to debate someone like Cybernetic Totalist philosopher Daniel Dennett. He would state that humans were simply specialized computers, and that imposing some fundamental ontological distinction between humans and computers was a sentimental waste of time.

"But don't you experience your life? Isn't experience something apart from what you could measure in a computer?", I would say. My debating opponent would typically say something like "Experience is just an illusion created because there is one part of a machine (you) that needs to create a model of the function of the rest of the machine- that part is your experiential center."

I would retort that experience is the only thing that isn't reduced by illusion. That even illusion is itself experience. A correlate, alas, is that experience is the very thing that can only be experienced. This lead me into the odd position of publicly wondering if some of my opponents simply lacked internal experience. (I once suggested that among all humanity, one could only definitively prove a lack of internal experience in certain professional philosophers.)

In truth, I think my perennial antagonists do have internal experience but choose not to admit it in public for a variety of reasons, most often because they enjoy annoying others.

Another motivation might be the "Campus Imperialism" I invoked earlier. Representatives of each academic discipline occasionally assert that they possess a most privileged viewpoint that somehow contains or subsumes the viewpoints of their rivals. Physicists were the alpha-academics for much of the twentieth century, though in recent decades "postmodern" humanities thinkers managed to stage something of a comeback, at least in their own minds. But technologists are the inevitable winners of this game, as they change the very components of our lives out from under us. It is tempting to many of them, apparently, to leverage this power to suggest that they also possess an ultimate understanding of reality, which is something quite apart from having tremendous influence on it.

Another avenue of explanation might be neo-Freudian, considering that the primary inventor of the idea of machine sentience, Alan Turing, was such a tortured soul. Turing died in an apparent suicide brought on by his having developed breasts as a result of enduring a hormonal regimen intended to reverse his homosexuality. It was during this tragic final period of his life that he argued passionately for machine sentience, and I have wondered whether he was engaging in a highly original new form of psychological escape and denial; running away from sexuality and mortality by becoming a computer.

At any rate, what is peculiar and revealing is that my cybernetic totalist friends confuse the viability of a perspective with its triumphant superiority. It is perfectly true that one can think of a person as a gene's way of propagating itself, as per Dawkins, or as a sexual organ used by machines to make more machines, as per McLuhan (as quoted in

the masthead of every issue of Wired Magazine), and indeed it can even be beautiful to think from these perspectives from time to time. As the anthropologist Steve Barnett pointed out, however, it would be just as reasonable to assert that "A person is shit's way of making more shit."

So let us pretend that the new Kant has already appeared and done his/her inevitable work. We can then say: The placement of one's circle of empathy is ultimately a matter of faith. We must accept the fact that we are forced to place the circle somewhere, and yet we cannot exclude extra-rational faith from our choice of where to place it.

My personal choice is to not place computers inside the circle. In this article I am stating some of my pragmatic, esthetic, and political reasons for this, though ultimately my decision rests on my particular faith. My position is unpopular and even resented in my professional and social environment.

Belief #4: That what Darwin described in biology, or something like it, is in fact also the singular, superior description of all possible creativity and culture.

Cybernetic totalists are obsessed with Darwin, for he described the closest thing we have to an algorithm for creativity. Darwin answers what would otherwise be a big hole in the Dogma: How will cybernetic systems be smart and creative enough to invent a post-human world? In order to embrace an eschatology in which the computers become smart as they become fast, some kind of Deus ex Machina must be invoked, and it has a beard.

Unfortunately, in the current climate I must take a moment to state that I am not a creationist. I am in this essay criticizing what I perceive to be intellectual laziness; a retreat from trying to understand problems and instead hope for software that evolves itself. I am not suggesting that Nature required some extra element beyond natural evolution to create people.

I also don't meant to imply that there is a completely unified block of people opposing me, all of whom think exactly the same thoughts. There are in fact numerous variations of Darwinian eschatology. Some of the most dramatic renditions have not come from scientists or engineers, but from writers such as Kevin Kelly and Robert Wright, who have become entranced with broadened interpretations of Darwin. In their works, reality is perceived as a big computer program running the Darwin algorithm, perhaps headed towards some sort of Destiny.

Many of my technical colleagues also see at least some form of a causal arrow in evolution pointing to an ever greater degree of a hard-to-characterize something as time passes. The words used to describe that something are themselves hard to define; It is said to include increased complexity, organization, and representation. To computer scientist Danny Hillis, people seem to have more of such a thing than, say, single cell organisms, and it is natural to wonder if perhaps there will someday be some new creatures with even more of it than is found in people. (And of course the future birth of

the new "more so" species is usually said to be related to computers.) Contrast this perspective with that of Stephen Jay Gould who argues in Full House that if there's an arrow in evolution, it's towards greater diversity over time, and we unlikely creatures known as humans, having arisen as one tiny manifestation of a massive, blind exploration of possible creatures, only imagine that the whole process was designed to lead to us.

There is no harder idea to test than an anthropic one, or its refutation. I'll admit that I tend to side with Gould on this one, but it is more important to point out an epistemological conundrum that should be considered by Darwinian eshatologists. If mankind is the measure of evolution thus far, then we will also be the measure of successor species that might be purported to be "more evolved" than us. We'll have to anthropomorphize in order to perceive this "greater than human" form of life, especially if it exists inside an information space such as the internet.

In other words, we'll be as reliable in assessing the status of the new super-beings as we are in assessing the traits of pet dogs in the present. We aren't up to the task. Before you tell me that it will be overwhelmingly obvious when the superintelligent new cyber-species arrives, visit a dog show. Or a gathering of people who believe they have been abducted by aliens in UFOs. People are demonstrably insane when it comes to assessing non-human sentience.

There is, however, no question that the movement to interpret Darwin more broadly, and in particular to bring him into psychology and the humanities has offered some luminous insights that will someday be part of an improved understanding of nature, including human nature. I enjoy this stream of thought on various levels. It's also, let's admit it, impossible for a computer scientist not to be flattered by works which place what is essentially a form of algorithmic computation at the center of reality, and these thinkers tend to be confident and crisp and to occasionally have new and good ideas.

And yet I think cybernetic totalist Darwinians are often brazenly incompetent at public discourse and may be in part responsible, however unintentionally, for inciting a resurgence of fundamentalist religious reaction against rational biology. They seem to come up with takes on Darwin that are calculated to not only antagonize, but alienate those who don't share their views. Declarations from the "nerdiest" of the evolutionary psychologists can be particularly irritating.

One example that comes to mind is the recent book, The Natural History of Rape by Randy Thornhill and Craig T. Palmer, declaring that rape is a "natural" way to spread genes around. We have seen all sorts of propositions tied to Darwin with a veneer of rationality. In fact you can argue almost any position using a Darwinian strategy.

For instance, Thornhill and Palmer go so far as to suggest that those who disagree with them are victims of evolutionary programming for the need to believe in a fictitious altruism in human nature. The authors say it is altruistic-seeming to not believe in evolutionary psychology, because such skepticism makes a public display of one's belief in brotherly love. Displays of altruism are said to be attractive, and therefore to improve one's ability to lure mates. By this logic, evolutionary psychologists should soon breed themselves out of the population. Unless they resort to rape.

At any rate, Darwin's idea of evolution was of a different order than scientific theories that had come before, for at least two reasons. The most obvious and explosive reason was that the subject matter was so close to home. It was a shock to the 19th century mind to think of animals as blood relatives, and that shock continues to this day.

The second reason is less often recognized. Darwin created a style of reduction that was based on emergent principles instead of underlying laws (though some recent speculative physics theories can have a Darwinian flavor). There isn't any evolutionary "force" analogous to, say, electromagnetism. Evolution is a principle that can be discerned as emerging in events, but it cannot be described precisely as a force that directs events. This is a subtle distinction. The story of each photon is the same, in a way that the story of each animal and plant is different. (Of course there are wonderful examples of precise, quantitative statements Darwinian theory and corresponding experiments, but these don't take place at anywhere close to the level of human experience, which is whole organisms that have complex behaviors in environments.) "Story" is the operative word. Evolutionary thought has almost always been applied to specific situations through stories.

A story, unlike a theory, invites embroidery and variation, and indeed stories gain their communicative power by resonance with more primal stories. It is possible to learn physics without inventing a narrative in one's head to give meaning to photons and black holes. But it seems that it is impossible to learn Darwinian evolution without also developing an internal narrative to relate it to other stories one knows. At least no public thinker on the subject seems to have confronted Darwin without building a bridge to personal value systems.

But beyond the question of subjective flavoring, there remains the problem of whether Darwin has explained enough. Is it not possible that there remains an as-yet unarticulated idea that explains aspects of achievement and creativity that Darwin does not?

For instance, is Darwinian-styled explanation sufficient to understand the process of rational thought? There are a plethora of recent theories in which the brain is said to produce random distributions of subconscious ideas that compete with one another until only the best one has survived, but do these theories really fit with what people do?

In nature, evolution appears to be brilliant at optimizing, but stupid at strategizing. (The mathematical image that expresses this idea is that "blind" evolution has enourmous trouble getting unstuck from a local minima in an energy landscape.) The classic question would be: How could evolution have made such marvelous feet, claws, fins, and paws, but have missed the wheel? There are plenty of environments in which creatures would benefit from wheels, so why haven't any appeared? Not even once? (A great long term art project for some rebellious kid in school now: Genetically engineer an animal with wheels! See if DNA can be made to do it.)

People came up with the wheel and numerous other useful inventions that seem to have eluded evolution. It is possible that the explanation is simply that hands had access to a different set of inventions than DNA, even though both were guided by similar processes. But it seems to me premature to treat such an interpretation as a certainty. Is it

not possible that in rational thought the brain does some as yet unarticulated thing that might have originated in a Darwinian process, but that cannot be explained by it?

The first two or three generations of artificial intelligence researchers took it as a given that blind evolution in itself couldn't be the whole of the story, and assumed that there were elements that distinguished human mentation from other Earthly processes. For instance, humans were thought by many to build abstract representations of the world in their minds, while the process of evolution needn't do that. Furthermore, these representations seemed to possess extraordinary qualities like the fearsome and perpetually elusive "common sense". After decades of failed attempts to build similar abstractions in computers, the field of AI gave up, but without admitting it. Surrender was couched as merely a series of tactical retreats. AI these days is often conceived as more of a craft than a branch of science or engineering. A great many practitioners I've spoken with lately hope to see software evolve that does various things but seem to have sunk to an almost "post-modern", or cynical lack of concern with understanding how these gizmos might actually work.

It is important to remember that craft-based cultures can come up with plenty of useful technologies, and that the motivation for our predecessors to embrace the Enlightenment and the ascent of rationality was not just to make more technologies more quickly. There was also the idea of Humanism, and a belief in the goodness of rational thinking and understanding. Are we really ready to abandon that?

Finally, there is an empirical point to be made: There has now been over a decade of work worldwide in Darwinian approaches to generating software, and while there have been some fascinating and impressive isolated results, and indeed I enjoy participating in such research, nothing has arisen from the work that would make software in general any better- as I'll ddescribe in the next section.

So, while I love Darwin, I won't count on him to write code.

Belief #5: That qualitative as well as quantitative aspects of information systems will be accelerated by Moore's Law.

The hardware side of computers keeps on getting better and cheaper at an exponential rate known by the moniker "Moore's Law". Every year and a half or so computation gets roughly twice as fast for a given cost. The implications of this are dizzying and so profound that they induce vertigo on first apprehension. What could a computer that was a million times faster than the one I am writing this text on be able to do? Would such a computer really be incapable of doing whatever it is my human brain does? The quantity of a "million" is not only too large to grasp intuitively, it is not even accessible experimentally for present purposes, so speculation is not irrational. What is stunning is to realize that many of us will find out the answer in our lifetimes, for such a computer might be a cheap consumer product in about, say 30 years.

This breathtaking vista must be starkly contrasted with the Great Shame of computer science, which is that we don't seem to be able to write software much better as computers get much faster. Computer software continues to disappoint. How I hated

UNIX back in the seventies - that devilish accumulator of data trash, obscurer of function, enemy of the user! If anyone had told me back then that getting back to embarrassingly primitive UNIX would be the great hope and investment obsession of the year 2000, merely because it's name was changed to LINUX and its source code was opened up again, I never would have had the stomach or the heart to continue in computer science.

If anything, there's a reverse Moore's Law observable in software: As processors become faster and memory becomes cheaper, software becomes correspondingly slower and more bloated, using up all available resources. Now I know I'm not being entirely fair here. We have better speech recognition and language translation than we used to, for example, and we are learning to run larger data bases and networks. But our core techniques and technologies for software simply haven't kept up with hardware. (Just as some newborn race of superintelligent robots are about to consume all humanity, our dear old species will likely be saved by a Windows crash. The poor robots will linger pathetically, begging us to reboot them, even though they'll know it would do no good.)

There are various reasons that software tends to be unwieldly, but a primary one is what I like to call "brittleness". Software breaks before it bends, so it demands perfection in a universe that prefers statistics. This in turn leads to all the pain of legacy/lock in, and other perversions. The distance between the ideal computers we imagine in our thought experiments and the real computers we know how to unleash on the world could not be more bitter.

It is the fetishizing of Moore's Law that seduces researchers into complacency. If you have an exponential force on your side, surely it will ace all challenges. Who cares about rational understanding when you can instead really on an exponential extra-human fetish? But processing power isn't the only thing that scales impressively; so do the problems that processors have to solve.

Here's an example I offer to non-technical people to illustrate this point. Ten years ago I had a laptop with an indexing program that let me search for files by content. In order to respond quickly enough when I performed a search, it went through all the files in advance and indexed them, just as search engines like Google index the internet today. The indexing process took about an hour.

Today I have a laptop that is hugely more capacious and faster in every dimension, as predicted by Moore's Law. However, I now have to let my indexing program run overnight to do its job. There are many other examples of computers seeming to get slower even though central processors are getting faster. Computer user interfaces tend to respond more slowly to user interface events, such as a keypress, than they did fifteen years ago, for instance. What's gone wrong?

The answer is complicated.

One part of the answer is fundamental. It turns out that when programs and datasets get bigger (and increasing storage and transmission capacities are driven by the same processes that drive Moore's exponential speedup), internal computational overhead often increases at a worse-than-linear rate. This is because of some nasty mathematical facts of life regarding algorithms. Making a problem twice as large usually

makes it take a lot more than twice as long to solve. Some algorithms are worse in this way than others, and one aspect of getting a solid undergraduate education in computer science is learning about them. Plenty of problems have overheads that scale even more steeply than Moore's Law. Surprisingly few of the most essential algorithms have overheads that scale at a merely linear rate.

But that's only the beginning of the story. It's also true that if different parts of a system scale at different rates, and that's usually the case, one part might be overwhelmed by the other. In the case of my indexing program, the size of hard disks actually grew faster than the speed of interfaces to them. Overhead costs can be amplified by such examples of "messy" scaling, in which one part of a system cannot keep up with another. A bottleneck then appears, rather like girdlock in a poorly designed roadway. And the backup that results is just as bad as a morning commute on a typically inadequate roadway system. And just as tricky and expensive to plan for and prevent. (Trips on Manhattan streets were faster a hundred years ago than they are today. Horses are faster than cars.)

And then we come to our old antagonist, brittleness. The larger a piece of computer software gets, the more it is likely to be dominated by some form of legacy code, and the more brutal becomes the overhead of addressing the endless examples of subtle incompatibility that inevitably arise between chunks of software originally created in different contexts.

And even beyond these effects, there are failings of human character that worsen the state of software, and many of these are systemic and might arise even if non-human agents were writing the code. For instance, it is very time-consuming and expensive to plan ahead to make the tasks of future programmers easier, so each programmer tends to choose strategies that worsen the effects of brittleness. The time crunch faced by programmers is driven by none other than Moore's Law, which motivates an ever-faster turnaround of software revisions to get at least some form of mileage out of increasing processor speeds. So the result is often software that gets less efficient in some ways even as processors become faster.

I see no evidence that Moore's Law is steep enough to outrun all these problems without additional unforeseen intellectual achievements.

A fundamental statement of the question I'm examining here is: Does software tend to be unwieldly only because on human error, or is the difficulty intrinsic to the nature of software itself. If there is any credibility at all to the eschatological scenarios of Kurtzweil, Drexler, Moravec, et al, then this is the single most important question related to the future of mankind.

There is at least some metaphorical support for the possibility that software unwieldliness is intrinsic. In order to examine this possibility I'll have to break my own rule and be a cybernetic totalist for a moment.

Nature might seem to be less brittle than digital software, but if species are thought of as "programs", then it looks like nature also has a software crisis. Evolution itself has evolved, introducing sex, for instance, but evolution has never found a way to be any speed but very slow. This might be at least in part because it takes a long time to

explore the space of possible variations of an exceedingly vast and complex causal system to find new configurations that are viable. Natural evolution's slowness as a medium of transformation is apparently systemic, rather than esulting from some inherent sluggishness in its component parts. On the contrary, adaptation is capable of achieving thrilling speed, in select circumstances. An example of fast change is the adaptation of germs to our efforts to eradicate them. Resistance to antibiotics is a notorious contemporary example of biological speed.

Both human-created software and natural selection seem to accrue hierarchies of layers that vary in their potential for speedy change. Slow-changing layers protect local theaters within which there is a potential for faster change. In computers, this is the divide between operating systems and applications, or between browsers and web pages. In biology, it might be seen, for example, in the divide between nature- and nurture-dominated dynamics in the human mind. But the lugubrious layers seem to usually define the overall character and potential of a system.

In the minds of some of my colleagues, all you have to do is identify one layer in a cybernetic system that's capable of fast change and then wait for Moore's Law to work it's magic. For instance, even if you're stuck with LINUX, you might implement a neural net program in it that eventually grows huge and fast enough (because of Moore's Law) to achieve a moment of insight and rewrite its own operating system. The problem is that in every example we know, a layer that can change fast also can't change very much. Germs can adopt to new drugs quickly, but would still take a very long time to evolve into Owls. This might be an inherent trade-off. For an example in the digital world, you can write a new JAVA applet pretty quickly, but it won't look very different from other quickly written applets- take a look at what's been done with applets and you'll see that this is true.

Now we finally come to...

Belief #6, the coming cybernetic cataclysm.

When a thoughtful person marvels at Moore's Law, there might be awe and there might be terror. One version of the terror was expressed recently by Bill Joy, in a cover story for Wired Magazine. Bill accepts the pronouncements of Ray Kurtzweil and others, who believe that Moore's Law will lead to autonomous machines, perhaps by the year 2020. That is the when computers will become, according to some estimates, about as powerful as human brains. (Not that anyone knows enough to really measure brains against computers yet. But for the sake of argument, let's suppose that the comparison is meaningful.) According to this scenario of the Terror, computers won't be stuck in boxes. They'll be more like robots, all connected together on the net, and they'll have a quite bag of tricks.

They'll be able to perform nano-manufacturing, for one thing. They'll quickly learn to reproduce and improve themselves. One fine day without warning, the new supermachines will brush humanity aside as casually as humans clear a forest for a new development. Or perhaps the machines will keep humans around to suffer the sort of indignity portrayed in the movie "The Matrix".

Even if the machines would otherwise choose to preserve their human progenitors, evil humans will be able to manipulate the machines to do vast harm to the rest of us. This is a different scenario that Bill also explores. Biotechnology will have advanced to the point that computer programs will be able to manipulate DNA as if it were Javascript. If computers can calculate the effects of drugs, genetic modifications, and other biological trickery, and if the tools to realize such tricks are cheap, then all it takes is a one madman to, say, create an epidemic targeted at a single race. Biotechnology without a strong, cheap information technology component would not be sufficiently potent to bring about this scenario. Rather, it is the ability of software running on fabulously fast computers to cheaply model and guide the manipulation of biology that is at the root of this variant of the Terror. I haven't been able to fully convey Bill's concerns in this brief account, but you get the idea.

My version of the Terror is different. We can already see how the biotechnology industry is setting itself up for decades of expensive software trouble. While there are all sorts of useful databases and modeling packages being developed by biotech firms and labs, they all exist in isolated developmental bubbles. Each such tool expects the world to conform to its requirements. Since the tools are so valuable, the world will do exactly that, but we should expect to see vast resources applied to the problem of getting data from bubble into another. There is no giant monolithic electronic brain being created with biological knowledge. There is instead a fractured mess of data and modeling fiefdoms. The medium for biological data transfer will continue to be sleep-deprived individual human researchers until some fabled future time when we know how to make software that is good at bridging bubbles on its own.

What is a long term future scenario like in which hardware keeps getting better and software remains mediocre? The great thing about crummy software is the amount of employment it generates. If Moore's Law is upheld for another twenty or thirty years, there will not only be a vast amount of computation going on Planet Earth, but also the maintenance of that computation will consume the efforts of almost every living person. We're talking about a planet of helpdesks.

I have argued elsewhere that this future would be a great thing, realizing the socialist dream of full employment by capitalist means. But let's consider the dark side.

Among the many processes that information systems make more efficient is the process of capitalism itself. A nearly friction-free economic environment allows fortunes to be accumulated in a few months instead of a few decades, but the individuals doing the accumulating are still living as long as they used to; longer, in fact. So those individuals who are good at getting rich have a chance to get richer before they die than their equally talented forebears.

There are two dangers in this. The smaller, more immediate danger is that young people acclimatized to a deliriously receptive economic environment might be emotionally wounded by what the rest of us would consider brief returns to normalcy. I do sometimes wonder if some of the students I work with who have gone on to dot com riches would be able to handle any financial frustration that lasted more than a few days without going into some sort of destructive depression or rage.

The greater danger is that the gulf between the richest and the rest could become transcendently grave. That is, even if we agree that a rising tide raises all ships, if the rate of the rising of the highest ships is greater than that of the lowest, they will become ever more separated. (And indeed, concentrations of wealth and poverty have increased during the Internet boom years in America.)

If Moore's Law or something like it is running the show, the scale of the separation could become astonishing. This is where my Terror resides, in considering the ultimate outcome of the increasing divide between the ultra-rich and the merely better off.

With the technologies that exist today, the wealthy and the rest aren't all that different; both bleed when pricked, for the classic example. But with the technology of the next twenty or thirty years they might become quite different indeed. Will the ultra-rich and the rest even be recognizable as the same species by the middle of the new century?

The possibilities that they will become essentially different species are so obvious and so terrifying that there is almost a banality in stating them. The rich could have their children made genetically more intelligent, beautiful, and joyous. Perhaps they could even be genetically disposed to have a superior capacity for empathy, but only to other people who meet some narrow range of criteria. Even stating these things seems beneath me, as if I were writing pulp science fiction, and yet the logic of the possibility is inescapable.

Let's explore just one possibility, for the sake of argument. One day the richest among us could turn nearly immortal, becoming virtual Gods to the rest of us. (An apparent lack of aging in both cell cultures and in whole organisms has been demonstrated in the laboratory.)

Let's not focus here on the fundamental questions of near immortality: whether it is moral or even desirable, or where one would find room if immortals insisted on continuing to have children. Let's instead focus on the question of whether immortality is likely to be expensive.

My guess is that immortality will be cheap if information technology gets much better, and expensive if software remains as crummy as it is.

I suspect that the hardware/software dichotomy will reappear in biotechnology, and indeed in other 21st century technologies. You can think of biotechnology as an attempt to make flesh into a computer, in the sense that biotechnology hopes to manage the processes of biology in ever greater detail, leading at some far horizon to perfect control. Likewise, nanotechnology hopes to do the same thing for materials science. If the body, and the material world at large become more manipulatable, more like a computer's memory, then the limiting factor will be the quality of the software that governs the manipulation.

Even though it's possible to program a computer to do virtually anything, we all know that's really not a sufficient description of computers. As I argued above: Getting computers to perform specific tasks of significant complexity in a reliable but modifiable

way, without crashes or security breaches, is essentially impossible. We can only approximate this goal, and only at great expense.

Likewise, one can hypothetically program DNA to make virtually any modification in a living thing, and yet designing a particular modification and vetting it thoroughly will likely remain immensely difficult. (And, as I argued above, that might be one reason why biological evolution has never found a way to be anything speed other than very slow.) Similarly, one can hypothetically use nanotechnology to make matter do almost anything conceivable, but it will probably turn out to be much harder than we now imagine to get it do any particular thing of complexity without disturbing side effects. Scenarios that predict that biotechnology and nanotechnology will be able to quickly and cheaply create startling new things under the sun also must imagine that computers will become semi-autonomous, superintelligent, virtuoso engineers. But computers will do no such thing if the last half century of progress in software can serve as a predictor of the next half century.

In other words, bad software will make biological hacks like near-immortality expensive instead of cheap in the future. Even if everything else gets cheaper, the information technology side of the effort will get more expensive.

Cheap near-immortality for everyone is a self-limiting proposition. There isn't enough room to accommodate such an adventure. Also, roughly speaking, if immortality was to become cheap, so would the horrific biological weapons of Bill's scenario. On the other hand, expensive near immortality is something the world could absorb, at least for a good long while, because there would be fewer people involved. Maybe they could even keep the effort quiet.

So, here is the irony. The very features of computers which drive us crazy today, and keep so many of us gainfully employed, are the best insurance our species has for long term survival as we explore the far reaches of technological possibility. On the other hand, those same annoying qualities are what could make the 21st century into a madhouse scripted by the fantasies and desperate aspirations of the super-rich.

Conclusion

I share the belief of my cybernetic totalist colleagues that there will be huge and sudden changes in the near future brought about by technology. The difference is that I believe that whatever happens will be the responsibility of individual people who do specific things. I think that treating technology as if it were autonomous is the ultimate self-fulfilling prophecy. There is no difference between machine autonomy and the abdication of human responsibility.

Let's take the "nanobots take over" scenario. It seems to me that the most likely scenarios involve either:

a) Super-nanobots everywhere that run old software- linux, say. This might be interesting. Good video games will be available, anyway.
b) Super-nanobots that evolve as fast as natural nanobots- so don't do much for millions of years.

c) Super-nanobots that do new things soon, but are dependent on humans. In all these cases humans will be in control, for better or for worse.

So, therefore, I'll worry about the future of human culture more than I'll worry about the gadgets. And what worries me about the "Young Turk" cultural temperament seen in cybernetic totalists is that they seem to not have been educated in the tradition of scientific skepticism. I understand why they are intoxicated. There IS a compelling simple logic behind their thinking and elegance in thought is infectious.

There is a real chance that evolutionary psychology, artificial intelligence, Moore's Law fetishizing, and the rest of the package, will catch on in a big way, as big as Freud or Marx did in their times. Or bigger, since these ideas might end up essentially built into the software that runs our society and our lives. If that happens, the ideology of cybernetic totalist intellectuals will be amplified from novelty into a force that could cause suffering for millions of people.

The greatest crime of Marxism wasn't simply that much of what it claimed was false, but that it claimed to be the sole and utterly complete path to understanding life and reality. Cybernetic eschatology shares with some of history's worst ideologies a doctrine of historical predestination. There is nothing more gray, stultifying, or dreary than a life lived inside the confines of a theory. Let us hope that the cybernetic totalists learn humility before their day in the sun arrives.

Reality Club Discussion

Margaret Wertheim

Science writer and commentator

I'd like to applaud Jaron's demi-manifesto. I heartily agree that what he called "cybernetic totalism" needs to be exposed. This indeed was one of the major themes of my own recent book The Pearly Gates of Cyberspace. I liked Jaron's analysis of what is wrong with cybernetic totalism very much, what was missing I think was an historical dimension as to why this way of thinking has evolved. Jaron rightly notes that this kind of thinking goes back to the dawn on the computer project with the work of Weiner and Shannon etc, but in fact this whole style of what I would label "techno-eschatology" has a much deeper history, going back to at least the middle Ages. Throughout Western history — since at least the twelfth century — there has been a very deeply ingrained tendency to link technology (in whatever is its recent mode) to an eschatological vision. Anyone interested in this subject should certainly read historian David Nobel's book The Religion of Technology, which traces the linking of technology to religious visions for the last millennium. In my own book I focus particularly on what might be briefly summarized as the religiosity inherent in our concepts of space, revealing the long historical roots of the belief in a transcendent "heavenly space" and the contemporary idea that cyberspace can be a new/ultimate realm of transcendence. Jaron is right that modern information theory has underlied the emergence of the belief that everything can be dissolved into information, but in parallel with this has also been a belief that beyond the mundane physical realm there exists an idealized "Platonic" realm of pure forms, pure data, pure

knowledge. This is also a critical dimension of cybernetic totalism, one which also has a long history in our culture.

What we need to understand, I suggest, is that the current iteration of techno-eschatology is nothing new in Western culture, that the techno/scientific culture of the West has indeed been pervaded by this spirit from the beginning. Which is not to say, of course, that all scientists and technologists think this way, only that there hasalways been a large contingent of our community who do. Like Jaron I believe we need to challenge this ideology — and it is an ideology — an especially pernicious one, I would argue. Like Jaron, I believe this ideology is crippling the advancement of science and technology (for this spirit inheres in much of the scientific community as well). It is also, as Jaron suggests, a force for exacerbating, not diminishing social equity. I am delighted to see this challenge being presented on the *Edge*, for on occasion I think that our community too has been too-heavily pervaded with a techno-eschatological spirit.

John Baez

Mathematical Physicist, U.C. Riverside

I found Jaron Lanier's half-manifesto very interesting. I doubt my friends in the academic humanities know people who are actively worrying about (or looking forward to) nanotech or a Vingean "singularity". They would probably dismiss such ideas as nuts. But as a scientist, I know quite a few such people: Extropeans, folks associated with the Foresight Institute, cypherpunks, fans of cryonics, and so on. So it makes sense to think seriously about what they are saying. I'm glad Jaron is doing this.

I don't have much to add except a couple of random remarks:

1) "The coming cybernetic cataclysm" takes various forms in the literature. Bill Joy's idea that autonomous machines will take over the world is actually a rather optimistic version. It assumes that machines, possibly with the help of "evil humans", will get good enough to beat us at our own game. My cybernetic totalist friends don't seem to worry about this scenario much. In fact, they may even relish the prospect! (Perhaps they are among those "evil humans" Joy talks about.)

What they worry about more is a "gray goo scenario" where due to some screwup, self-replicating unintelligent nanotech gets loose which eats the entire biosphere. Myself, I'm not sure if this a paranoid fantasy or a realistic possibility, since I don't know whether the biosphere is operating near maximal efficiency or whether a small, simple new entity could manage to eat everything in its path without being eaten itself. But in general, I'm more afraid of stupid mistakes than an attack of superintelligent beings. The gray goo scenario is just one of many possible mistakes we might make. The really dangerous ones are the ones we won't think of until they're already happened.

2) I liked Jaron's remark about physicists being the "alpha-academics" for most of the last century. It's curious being a mathematical physicist now that this era is over. I went into this field as a kid because I thought it was the coolest thing around. Gradually I realized that it's not — at least, not as measured by the standards of money and power! At first this was a bit of a letdown, but now it seems liberating in some respects. I don't have to worry that my research on quantum gravity will be used to create a super-bomb or destroy the universe — at least not in the near future — because we have no way of

accessing the energy scales needed to wreak havoc in that way. Besides, we can blow ourselves up quite nicely already. Now it's the computer science, biotech and nanotech people who have to shoulder the responsibility of doing science that seriously affects human lives, while I enjoy playing around with my equations.

Yes, I'm being a bit sarcastic here, but it is very interesting how these things change.

Lee Smolin

Physicist, Perimeter Institute; Author, Einstein's Unfinished Revolution

Jaron is raising some very important points that deserve closer examination and discussion. Among them is his challenge to the idea that the optimization of present day computers could produce anything with the capabilities of living, intelligent animals, cats let alone people. I think Jaron is right to point out that the arguments for this thesis rest on incorrect assumptions. I believe that Jaron's argument can be strengthened and I would like to explain how. The following is just a sketch, but I hope it suffices to stimulate the debate.

The problems to be addressed are 1) what kinds of problems can computers solve and whether they differ in kind from the kinds of problems humans solve. 2) What kind of problem is it to design a computer and whether it differs in kind from the problem of designing a human, or a creature with equal capabilities.

To approach these questions it helps to begin with the idea that some design problems involve searching a space of possible design parameters. We know that in these cases there are simple optimization algorithms that will find the local extrema in whatever basin of attraction one happens to be in. However, optimization is a small part of design because it can be used reliably to solve only a small subset of possible design problems. To talk about this we may distinguish five classes of design problems.

CLASS 1: Local optimization problems problems which can be solved with standard hill-climbing techniques.

CLASS 2: Locate a pretty good, but not necessarily global extremum in a configuration space with many local extrema and many basins of attraction.

CLASS 3: Locate the global extremum in a configuration space with many local extrema and many basins of attraction.

CLASS 4: Find local extrema in a landscape which changes unpredictably on the same time scale it takes to find local optima.

CLASS 5: find local extrema in cases in which the computation time required to construct the configuration space and/or calculate the fitness function is either infinite or much longer than the time available. These are the class of problems which have to be invented or discovered before they can be solved, as there is no algorithm that can lead to their formulation or complete specification.

Let us first discuss the first question. At least so far, computers are very good at solving CLASS 1 problems, and there are decent algorithms for simple CLASS 2 problems. But we do not have good methods for finding global extrema and hence solving CLASS 3 problems. To my knowledge computers can do decently at some simple CLASS 4 problems, but can easily fail when they become more complex. By definition

computers have problems solving CLASS 5 problems, as the computation time to set up the extremization problem is prohibitive. However humans can often solve CLASS 3 problems and are also quite good at CLASS 4 problems. This should be no surprise, this is part of our biological specialization. This is what is required to flourish in a new environment, domesticate a new species, become farmers, populate almost all the ecological zones on the planet and so forth.

But humans can do even better than that, we can both invent and solve CLASS 5 problems. This is what poetry, art, music and science, are about. We invent the forms and traditions and then we master them. We can thrive in a domain in which we create optimal versions of things that did not even exist a short time before. We are not extremizing in a landscape, we are building the landscape on the same time scale that we master it.

One correspondent suggested that anyone who thinks people are different from machines are naive romantics. This is not true, we are different because we have vastly different capabilities. It is irrelevant to talk of the universality of Turing machines, for Turing machines are entities that run programs that must be written by an external entity. So far at least the only entities we know of who can function as those external programers are humans. Humans are intelligent creatures that do not need to be programmed by any external agency. Turing machines are designed, we are the result of natural selection. We need then to examine the second question, whether designing or programming a computer is in the same CLASS of problems as the problems natural selection solved in the course of evolution.

Of course inventing the idea of a digital computer was a CLASS 5 problem. But once we had the idea, the optimization of digital computers is mainly a CLASS 1 problems. This is what Moore's law is about, it tells us how quickly local optimization can work when ample resources are available. One of the points Jaron is making is that the design of software required to do justice to the exponentially increasing capabilities of our machines are not CLASS 1 problems. Moore's law tells us that the fitness landscape for software is changing on a time scale comparable to the time required to write and debug software. Thus writing software involves problems of at least CLASS 4. This is of course just a different way of making one of Jaron's arguments.

For there to be a danger of robots taking over, or even being able to do a decent job entertaining us, replacing songwriters and singers,artists, scientists and comedians, one of two things have to happen. Either we will be able to design a machine that could replace us, which means a machine that can solve problems of CLASS 5, or we will be able to design a machine that could in turn design a machine that could solve CLASS 5 problems.

But while we can solve problems up to CLASS 5, so far we have only been able to design machines that can solve CLASS 2 problems reliably. And so far machines are not able to design other machines to solve even CLASS 1 problems. When one puts it this way it is clear that it is not just a matter of Moore's law, designing one of us is a very different kind of problem then optimizing a programmable digital computer.

What kind of problem is it to design an entity that can solve CLASS 5 problems? We know we were created by natural selection, acting on not only us but the whole collection of living species. This is at least a CLASS 4 problem, but it is very likely at least a CLASS 5 problem. The interactions among many species as they evolve under the rules of natural selection is a CLASS 4 problem, as is shown by models of Bak and Sneppen, Kauffman, Sola and others. But there are good arguments, summarized in Stuart Kauffman's forthcoming book, that natural selection and cultural evolution are really CLASS 5 problems. He argues that they are problems in which the construction of the fitness landscape itself is so computationally intensive that it is not correct to separate the specification of the fitness landscape from its optimization. Instead, both take place together. This means really that the metaphor of optimization has broken down completely. Whatever evolution is doing cannot, he argues, be conceptualized as extremization on a pre-existing fitness landscape.

Thus, the problem of designing an entity that can solve CLASS 5 problems is at least a CLASS 4 problem, and very likely is a CLASS 5 problem. But is it only this hard, or harder still? Human's can solve some CLASS 4 and 5 problems, but it is not at all obvious that the problems of these kinds that we can solve are comparable to the problems that natural selection has solved in designing us. At the very least, it is likely that the time required to solve the problem of designing us may take a great deal longer than the tine it takes to solve the CLASS 4 and 5 problems we have so far dealt with. It took natural selection 4 billion years to design us. Let us assume that we could do it much faster. How much faster? Let us assume that we could use genetic engineering to engineer an artificial speciation in an animal. Speciation is a process that takes on the order of 100,000 years. Given very optimistic assumptions it is possible to imagine that some years from now this is something we will be able to accomplish in on the order of 100 years. It could certainly not be less than that as we cannot do it faster than the time it takes for several generations to grow to maturity. (Because the interaction of an animal and its environment is a CLASS 5 problem, we are not likely to be able to simulate it reliably enough to replace the phase where we grow the animal and observe what happens.) This would mean that we had the tools to speed up natural selection by a factor of 1,000. Even with this fantastic increase of speed it would still take us a million years to invent something like ourselves, starting from scratch. (Note that this is true even if we skip the pre cambrian stages of evolution, which begins with creatures whose cell biology and biochemistry is far advanced of what we have so far designed. Note also that many biologists working in parallel won't help as natural selection also works in parallel.)

This is on the order of the lifetime of a species. A problem like this, whose minimum time for solution is on the order of the lifetime of a whole species of creatures that can solve CLASS 5 problems deserves a separate class. So we may call this a CLASS 6 problem.

Is it possible that there is a way to do it much faster, by taking a route that natural selection could not have? One cannot say this is impossible, but all this means is that so little is known about the problem that it is in a class of problems we have no idea how to solve.

To summarize: the claim that optimization of present computer designs could produce something that is "as powerful" as humans requires that there is only one kind of intelligent entity, and they all live in a in a fixed landscape with a single local extremum. But we are not only not in the same basin of attraction as present day computers, it is not even obvious that the problem of constructing us has anything in common with problems we have so far solved. This is not to deny that someday humans may learn how to solve the problem of designing creatures that can themselves solve CLASS 5 problems. The point is only that there is no rational basis for predicting when or even whether this may happen, as the solution to this problem is not closely related to the kind of optimization problems that human designers have so far learned to solve.

Stewart Brand

Founder, the Whole Earth Catalog; Co-founder, The Well; Co-Founder, The Long Now Foundation, and Revive & Restore; Author, Whole Earth Discipline

What a juicy piece of work by Jaron!

For me, one ancillary proof of much of his thesis is the phenomenon of Libertarian politics, which I've considered to be algorithmic political pseudoscience and now, thanks to Jaron, consider to be an offshoot of Cybernetic Totalism. Libertarian thinking is a common (though certainly not universal) affliction of working computeroids and their followers. Struck dumb by the cybernetic marvel of the marketplace, with its self-balancing and even fractal Invisible Hand, Libertarians seem unwilling to consider the equally marvelous cybernetic structure of the US Constitution or to consider that the sheer messiness of democracy in action is part of the system's long-term health.

Libertarians get caught up in simplistic analyses such as that since police departments require crime in order to exist, therefore they are incented to make sure that crime is never "solved," creating it themselves if necessary. Or, more subtly, that since competition forces competitors to become more alike, therefore police will become like criminals so much that they are, in fact, criminals after a while. Both ideas are helpful, but there is no place in such analyses for trans-logical concepts like "honor" or "service," and they are what drive a huge part of effective police work.

George Dyson

Science Historian; Author, Analogia

Without taking one side of Jaron's dogma or another (place me somewhere else entirely) I would disagree strongly with his "Argument from Software" — which is as flawed as Bishop Wilberforce's Argument from Design.

Back in the days when programs could be debugged but processing could not be counted on from one kilocycle to the next, John von Neumann wrote his final paper in computer theory: "Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components" [in Claude Shannon and John McCarthy, eds., Automata Studies(1956) pp. 43 — 99]. It makes no difference whether you have reliable code running on lousy hardware, or lousy code running on reliable hardware. Same results.

What should reassure the technophiles, and unsettle the technophobes, is our world of lousy code. Because it is lousy code that is bringing the digital universe to life, rather than leaving us stuck in some programmed, deterministic universe devoid of life. It

is that primordial soup of archaic subroutines, ambiguous DLL's, crashing Windows, and living — fossil operating systems that is driving the push towards the sort of fault embracing template — based addressing that proved so successful in molecular biology, with us — and our computers — as one of its strangest results.

Let us praise sloppy instructions, as we also praise the Lord.

Rodney A. Brooks

Panasonic Professor of Robotics (emeritus); Former Director, MIT Computer Science and Artificial Intelligence Lab (1997-2007); Founder, CTO, Robust.AI; Author, Flesh and Machines

Lee Smolin wrote:

"One correspondent suggested that anyone who thinks people are different from machines are naive romantics. This is not true, we are different because we have vastly different capabilities. It is irrelevant to talk of the universality of Turing machines, for Turing machines are entities that run programs that must be written by an external entity."

This is exactly the sort of naive romanticism to which I was refering. I was not comparing humans to a PC running Windows 2000. I am saying that people are machines in the sense that there is, as far as we have any scientific knowledge at this time, nothing in them outside the laws of physics of the universe which govern all matter. People are made of matter and that matter obeys the physical laws of the universe. Unless one hypothesizes an eternal soul, an elixir of life, an ineffable essence, or some other extra-physicalness to humans (and also to other animals, all the way down to bacteria?), then humans are machines. It has absolutely nothing to do with Turing machines, or programming computers.

Get over your fear of being a machine. We are not the center of the universe, and God does not exist. That is what this disagreement boils down to.

Freeman Dyson

Physicist, Institute of Advanced Study; Author, Disturbing the Universe; Maker of Patterns

Dear George, your reply to Lanier is brilliant, profound, and also true. I remember that I wrote, at the end of Origins of Life, that the evolution of complex organisms became possible when the essential sloppiness and error — tolerance of life were transferred from the hardware to the software, from the metabolic apparatus to the genes. And now you are saying that exactly the same thing happened in the evolution of complex computer — systems. Obviously, that's the direction you have to go if you want to combine robustness with creativity. All I can say is, why didn't I think of that?

Lee Smolin

Physicist, Perimeter Institute; Author, Einstein's Unfinished Revolution

In reply to Rodney Brooks:

I believe strongly that our entire existence is as part of the natural world. I am not afraid of this; my book, The Life of the Cosmos, is a kind of homage to that idea. My guess is that we agree broadly on metaphysics, but my comment had nothing to do with

God, cosmology, consciousness or any kind of romanticism. I was trying to make a point about science, one that is well within the boundaries of our shared metaphysics.

In my comment I raised two issues. First, whether everything that is part of the physical universe can be described in terms of a Turing machine, second, whether the way that living animals process information is enough like how digital computers work that it is rational to hope to construct a reasoning animal based on models of digital computers. As these seem to be very open issues given the present evidence, it seems far from clear that the metaphor of a machine will in the end be very helpful to us as in understanding in physical terms what animals are. In addition there is a problem with using the word machine in this context, which is that it carries with it the implication that something was made by human beings. This is not just semantics because ignoring the deep differences-as physical systems-between living animals and human made machines has led to some predictions for the future of machines that may not be consistent with our developing understanding of what life is.

To expand on this last point, I do believe that we will someday understand what we are in terms of physics. But before we do that we must first understand what a living thing is in terms of the laws of physics. We have made a lot of progress towards this in the last years and I believe more will be made shortly. Everything we have learned suggests that there are important differences, expressible in completely physical terms-more particularly in terms of statistical physics, between systems that are made and systems systems that arise by a spontaneous process of self-organization. Both may process information, but they may do so in different ways, so that they are generally able to solve different classes of problems.

A related point is made by Stuart Kauffman in recent papers and a forthcoming book: there is a fundamental difference between a physical system that can be termed an "autonomous agent" and one that cannot be. Part of Kauffman's definition of an autonomous agent is that it is a self-reproducing system, able to carry out at least one thermodynamic work cycle. Computers are not autonomous agents to the extent that they are constructed and programmed. But computers are Turing machines-which is why that idea is useful for this discussion.

Living animals are autonomous agents. They are not, so far as has been shown, Turning machines. There is no obvious relationship between the definition of a Turing machine and the definition of an autonomous agent; it is certainly very unlikely that they are equivalent. Thus, while it is of course possible that we may some day be able ourselves to make living things, there does not seem to be any good reason to expect that such articial animals will have a strong resemblance structurally or functionally to computers. (The fact that one can model certain aspects of life in computer software does not change this.)

Computers are wonderful tools and fantastic toys. But if machine is to mean anything at all besides "something found in the universe" (remember that we have the same metaphysics) then computers are machines, and animals are not.

Cliff Barney

Former Journalist

Jaron Lanier argues persuasively, but in a social vacuum. Cyberarmageddon, feasible or not by 2020, would be not a technological but a social phenomenon. Lanier argues that it won't happen because it can't, computers being what they are; possibly true but irrelevant to the great social upheavals that are occurring today in 2000 as information technology develops. These changes, as much as Moore's Law, will determine how technology develops. "Society doesn't work technologically," says Manuel Castells; "technology is used and reused and adapted by society."

The Dalai Lama put it another way: "Technology is not the basis of our society, compassion is the basis of society."

We do in fact have one million — fold increase in computer power to look at: the jump between 1968, when Doug Engelbart invented the mouse, and 1998, when he and his many friends lamented at Stanford that it hadn't changed the world as much as they imagined it would (see www.netfront.to/Engel1.html). Thirty years, doubling every year and a half, gives us the millionfold power increase, and we went from mainframes with bales of wire hanging out the back to palmtops and satellites. What were the social changes?

Castells has catalogued these, and offered a hypotheses for understanding the change, in his three — volume survey of The Information Age, in which he describes the Network Society that has emerged in the past 30 years. We can see some of the results in the post — Seattle streets, as individuals attempt to find an identity vis — ‹ — vis a global economic network. This is Moore's Law, social edition.

In this respect I wish Lanier had written the other half of his manifesto — the part about the "lovely global flowering of computer culture already in place." This is more likely to affect Armageddon than Dr. Moore's relentlessly shrinking etchings.

Daniel C. Dennett

Philosopher; Austin B. Fletcher Professor of Philosophy, Co-Director, Center for Cognitive Studies, Tufts University; Author, From Bacteria to Bach and Back

A friendly alert to Jaron Lanier

Unalloyed enthusiasm for anything is bound to be a mistake, so thank goodness for the critics, the skeptics, the second-thought-havers, and even the outright apostates. Apparently the price one must pay for jumping off a fast moving bandwagon is missing the target somewhat, since it seems that apostates usually overstate the case and land somewhere rather far from where they aimed. Reading Jaron Lanier's half a manifesto, I was reminded of an earlier critic of digital dreams, Joseph Weizenbaum, whose 1976 book, Computer Power and Human Reason, was an uneven mix of serious criticism in the tradition of Norbert Wiener and ill-developed jeremiads. Weizenbaum, in spite of my efforts (for which I was fulsomely thanked in his preface), could never figure out if he was trying to say that AI was impossible, or all-too-possible but evil. Was AI something we couldn't develop or shouldn't develop? Entirely different cases, requiring different arguments. There is a similar tension in Lanier's writing: are the Cybernetic Totalists just hopelessly wrong—their dream is, for deep reasons, impossible—or are they cheerleaders we must not follow—because we/they might succeed? There is an interesting middle course, combining both options in a coherent possibility, and I take it that this is the best

reading of Lanier's manifesto: the Cybernetic Totalists are wrong and if we take them seriously we will end up creating something—not what they dream of, but something else—that is evil.

But who are the Cybernetic Totalists? I'm glad that Lanier entertains the hunch that Dawkins and I (and Hofstadter and others) "see some flaw in logic that insulates [our] thinking from the eschatalogical implications" drawn by Kurzweil and Moravec. He's right. I, for one, do see such a flaw, and I expect Dawkins and Hofstadter would say the same. My reason has always been that the visionaries who imagine self-reproducing robots taking over in the near future have bizarrely underestimated the complexities of life. Consider the parallel flaw in the following passage from truth to foolishness:

TRUE: living bodies are made up of nothing but millions of varieties of organic molecules organized by the trillions into complex dynamic structures such as cells and larger assemblies (there is no élan vital, in other words).

FOOLISH CONCLUSION: therefore we shall soon achieve immortality; all we have to do is direct all our research and development into molecular biology with the goal of replacing those individual molecules, one at a time, as they break or wear out.

You don't have to be a vitalist to reject this technocratic fantasy, and you don't have to be a dualist, an anti-mechanist, to reject simplistic visions of some AI utopia just around the corner. Lanier is wistful about the possibility "that in rational thought the brain does some as yet unarticulated thing that might have originated in a Darwinian process, but that cannot be explained by it [my italics]," but why should it matter? Lanier is too clever to ask for a skyhook, but he can't keep himself from yearning for . . . . half a skyhook.

It is ironic that when Lanier succumbs to temptation and indulges in a bit of cybernetic totalism of his own, he's pretty good at it. His speculative analysis of the inevitability of what might be called legacy inertia, creating diminishing returns that will always blunt Moore's law, is insightful, and I welcome these new reasons his essay gives me for my skepticism about the cybernetic future. But I wish he didn't also indulge in so much presumptive caricature of those positions he finds threatening. He apparently doesn't want there to be subtle, nuanced, modest versions of the theses he resists, since those would be so hard to sweep away, so he follows the example of one of his heroes, Stephen Jay Gould, and stoops to the demagogic stunt of creating strawpeople and then blasting away at them. He's got me wrong, and Dawkins, and Thornhill and Palmer, to name the most obvious cases. It's child's play to hoot at parodies of me on consciousness, Dawkins on memes, Thornhill and Palmer on rape. Grow up and do some real criticism, worth responding to. We're not the bad guys; we hold positions that are entirely congenial to his trenchant criticisms of simplistic thinking about computation and evolution.

Joseph Weizenbaum soon found himself drowning under a wave of fans, the darling of a sloppy-thinking gaggle of Euro-intellectuals who struck fashionable Luddite poses while comprehending almost nothing about the technology engulfing them. Weizenbaum had important, reasoned criticisms to offer, but all they heard was a Voice

on Our Side against the Godless Machines. Jaron, these folks will love your message, but they are not your friends. Aren't your criticisms worthy of the attention of people who actually will try to understand them?

Bruce Sterling

Science Fiction Author, Mirrorshades

Jaron has written a very beautiful work. This screed is truly a native document of the year 2000 AD. I felt very privileged and happy to read this. It really floods the mind with its clarity and insight. It's very musical.

I've been thinking a long time about the "eschatological cataclysm" detailed in "Belief #6." This is known in my trade as the "Vingean Singularity," and us undignified pulp science fiction writers consider it particularly galling, because this is a point at which our craft breaks down. It's a Lanierian software traffic jam for science fiction, really, where our ability to generate and scatter mindblowing concepts outruns the inherent limitations of our merely human frontal lobes.

I have now rubbed — up against the stark cosmic horror of Belief #6 long enough to get rather chummy and cozy with it, and would like to offer a new, brief set of corollary beliefs.

1. There is no one Singularity. Any area of scientific inquiry, pushed far enough, could provide its own native version of a cataclysm: biological, cognitive, mechanical, cybernetic, you could name it. If man is the measure of all things, then there probably is no measure by which we can't be made more than human.

2. A Singularity ends the human condition (because that is its definition), but it resolves nothing else. It would almost certainly be followed by a rapid, massive explosion of following Singularities. These ultra — cataclysmic events would disrupt the first Singularity even more than the first Singularity disrupted the human condition.

3. The posthuman condition is banal. It is crypto — theological, and astounding, and apocalyptic, and eschatological, and ontological, but only by human standards. Oh sure, we become as gods (or something does), but the thrill fades fast, because that thrill is merely human and parochial. By the new, post Singularity standards, posthumans are just as bored and frustrated as humans ever were. They are not magic, they are still quotidian entities in a gritty, rules — based physical universe. They will find themselves swiftly and bruisingly brought up against the limits of their own conditions, whatever those limits and conditions may be.

4. Messy, embarrassing, reversible, goofy, catch — as — catch — can posthumanism is politically preferable to sleek, streamlined, sudden, utter, Final Solution posthumanism. The best way to encounter a Singularity would be to nick over the event horizon for a minute or two and have somebody else yank you back. Then the rest of us would be able to debrief you, and see if you could still write as well as Jaron Lanier.

Philip W. Anderson

Nobel Laureate; Physicist

I was very happy to see Jaron Lanier's paper, in that it was saying a lot of things I had felt to be true, and saying them from within the digital world. The twenty-year

prediction for conscious robots reminds me, for instance, of the twenty years since Stephen Hawking's prediction that in the year 2000 there would be no more theoretical physicists, only computers. What has actually happened has been that the currently fashionable field in theoretical physics, superstring theory, is an almost entirely analytical development. Computers can't even yet do respectable field theory for simple systems in four dimensions, much less 10 or more. What is happening in the rest of theoretical physics is even more depressing — if you find that depressing, that is — which is that the government agencies have been sold a bill of goods by you digerati, and will fund happily only theoretical physics done by computer; whereas the real problems are those which have not yet been conceptualised and simplified enough to use a computer.

I ran across a quote from, oddly, G. K. Chesterton, which makes one of the points nicely. "life is a trap for logicians; it looks just a little more mathematical and regular than it is. Its exactitude is obvious, but its inexactitude is hidden; its wildness lies in wait."

I also felt a resonance with another story. I read recently a review of a book in which it is shown that the Alexandrians had reached a level of scientific sophistication by 150 BC which was close to that of 17th century England; for instance, that much of Newton's Principia borrowed ideas from Greek texts. The science was lost, he claimed, not by the Church, although it sure helped, but by the practical engineering bent of the Romans, who took the useful engineering rules of thumb like Ptolemy's cycles and ignored the science behind them. In other words, exactly the "dumbing down" process that Lanier describes.

I am also fascinated by Lanier's idea that there is something between simple digital representations of input data and the "qualia" of a dualistic animalcule. The architecture of the brain attaches a very complex structure to each region of the visual or tactile field, a kind of a minibrain connected to all the other minibrains. Presumably this minibrain doesn't tell all the others that its part of the field has thus-and-so spectrum, it tells them that it's red (or at least redder than their part).

In general, I think there is much too much of a tendency to think that a representation of the world in terms of bit strings is a satisfactory one (even if complete). If this is so, why does the quantum computer do new things? Why is complexity theory such a poor guide to the real world of problems?

A decade ago I reviewed a book about Ed Fredkin (among others) in which he expressed the opinion that even the ultimate fine structure of space-time was digital. This bad idea was later taken up by John Wheeler (he calls it "it from bit") as well as a number of other less able physicists. The problem with it is that all of our success with particle physics — the Standard Model — is based upon continuous symmetries to which a digital picture is maximally unsuited. Modern quantum gravity actually claims to be seeing the scale at which it all stops, and if you can believe their picture it sure doesn't look digital at all. (They describe it as all the theories seguing into each other, kind of, but none of them are discrete.)

I guess the problem I have is that discrete mathematics feels too anthropomorphic — too much creating the world in our own image. No matter how far Moore's law carries

us, it is still digital. I am not agreeing with Penrose, nor do I believe we are anything but a machine — but are we a digital machine? To put it less mystically, is a digital representation practical?

Rodney A. Brooks

Panasonic Professor of Robotics (emeritus); Former Director, MIT Computer Science and Artificial Intelligence Lab (1997-2007); Founder, CTO, Robust.AI; Author, Flesh and Machines

I do not at all agree with Moravec and Kurzweil's predictions for an eschatological cataclysm, just in time for their own memories and thoughts and personhood to be preserved before they might otherwise die. I do not discount that the logical consequence of some version of cybernetic totalism might ultimately happen. I just happen to think it is going to be somewhat different in form than the version discussed by Lanier, and probably will happen much more gradually over some centuries, with no visible cataclysm, and no real eschatological division between before and after. And, I agree entirely with Lanier that the particular arguments of Moravce and Kurzweil seem to rely too much on it all happening just because there will be lots of more of Moore's law computer power. Neither Moravec or Kurzweil ever give a hint of what technical innovations need to be made to get to intelligent machines that will be able to do all the things they predict. Lanier however seems to deny that any such thing could ever happen, and his arguments largely boil down to an inbuilt fear of losing a last bastion of human specialness.

But first let me complain about one particular technical view expressed in Lanier's manifesto. Occasionally emerging is a fear of nanotechnology. It is not clear exactly which version of nanotechnology he fears, but I have become increasingly annoyed at the hyping up of concerns about "strong" nanotechnology that Bill Joy and others have recently engaged in. Lanier's super — nanobots in his conclusion certainly smack of strong nanotechnology. Strong nanotechnology, the version that is most popular in science fiction, has molecular machines which can manipulate matter, disassemble arbitary raw materials atom by atom, and build copies of themselves. We do not know whether the physics of our universe allows such machines to exist, or whether self reproducing machines need to use the molecular mechanisms of biology and must be on the order of billions of atoms in size. What we have seen so far in nanotechnology is the ability for us to manipulate single atoms in carefully controlled conditions using multi — kilogram machines. We have no evidence that non — biological nanotechnology machines will even in principle be able to manage energy supplies, manipulate single atoms in arbitrary ways, break down raw materials, both decode and copy a description of themselves, implement the computational resources necessary to control their behavior, and avoid being ripped asunder by the presence of other nearby matter. We have no clue when we will be able to answer whether such machines can exist, even in principle. Worrying about whether nanotechnology machines might "get away" from us and eat the fabric of our world, or evolve to do so, seems to me to be on a par with worrying about how the world will fare with the screwups in temporal consistency that will occur once

we have figured out how to build time travel machines. Another topic popular in science fiction.

Now to the main disagreement I have with Lanier.

The first problem I have is with his dismissal of Artificial Intelligence as being based on an intellectual mistake. His argument is all smoke and mirrors with no viable logic. He uses the Turing test as the touchstone for AI, and argues that besides the computer getting as smart as a person, the Turing test could also be passed by a computer if the people get dumber. He claims the second is happening, and with a flourish worthy of a stage magician draws attention away from the first possibility, in effect negating that it might ever happen, just because he has anecdotal frustrations with business software systems illustrating cases of the second. This is no argument!

Then we get to Lanier's real failure. He turns out to be a closet Searlean. He "experiences" life, and no computer, he implicity argues, can every "experience" life. Why not? More smoke and mirrors...he has talked to philosophers who do not tackle his argument head on. If we accept that living systems are made up of physical molecules, and nothing non — tangible external to an understanding of the physics of the world, no essence, no immortal soul, no elixer of life, then we humans are machines and we humans do "experience" life. I do. A lot of the time. I see no reason therefore that other machines, that don't happen to have the same biological history as me can not also "experience" life. Searle argues that an atom for atom reproduction of me will act like me by will not really "experience" life. Lanier does not get into this level of detail, but clearly he (and Searle) and I have different dogmatic understandings of the universe. He requires some implicit specialness for biological people, I require that in principle non — biological machines can "experience" life. I do not quite know how to build such machines yet in detail, but it is perhaps no more of a stretch to have explained the heart as a pump delivering oxygenated blood to the body before the structure of hemoglobin was understood. The explanation certainly semed right, even before the details were known.

Mankind, and probably Lanier, has had to give up the notion that the earth is special and the center of the universe, has had to give up the notion that god created animals and humans in fundametally different ways but instead both were produced by evolution and natural selection, and has had to give up the notion that we are vastly different from yeast in our fundamental biochemical pathways. What is left for us proud humans is that we are different from machines in some fundamental ineffable way. Lanier does not want to give this up. I am willing to.

I'll take the null hypothesis. We are machines until proven otherwise, rather than just wished otherwise. Whether people are smart enough to build machines that "experiences" the world is another question. But in principle it can surely be done, and hence the cybernetic totalism that Lanier so irrationally, and tribally, fears.

Jaron Lanier

Computer Scientist; Musician; Author, Who Owns The Future?

Hello to two generations of Dysons, Freeman and George, both of whom I admire. I must say that it is immediately apparent that our priorities are different. As I hope my essay makes clear, I am more concerned with how people design technology

and relate to it psychologically than with the long term fate of the machines themselves. Whether or not George Dysons' critique is technically correct, in my opinion it is esthetically, ethically, and politically misguided, in that he is looking at questions solely from the perspective of the machines rather than from the perspective of people. I see that I have genuinely failed to communicate this most essential point in my essay across a cultural chasm, and it saddens me. My failure is made more plain by the flip theological references in the George Dyson's note; he is apparently more comfortable deifying software than in recognizing the value of human aspirations to rational design.

If a future develops in which Dyson would perceive new life forms to have arisen from adaptations of messy software, I would perceive instead a lot of anti-human programming and design resulting in opaque user interfaces, i.e. machines that no longer made sense to people. I would also perceive a loss of human drive to achieve elegance in software design and an abandonment of rational planning. The most important point in my essay is that our two differing interpretations would each be reasonably applicable to the same outcome. I am advocating one interpretation over the other for reasons that arise from human, rather than technical concerns.

The argument that the Dysons do address is a secondary one in my mind; to what degree messiness limits or enhances the future of software. The key question here is whether different kinds of unreliability are effectively interchangeabled. George Dyson equates the failure modes of primordial chemistry with failure modes seen in contemporary software. This shouldn't be understood as a comparison between hardware and software per se, but between elements whose connections can only be described by statistics, like molecules, or indeed physical gates in a computer, versus elements that connect by Platonic logic.

Certainly the Dysons are correct to a degree, in the sense that error recovery algorithms can grant a "soft knee" to software failure modes that is reminiscent of the type of "statistical binding" seen in natural systems. Real computers as we know them are not built this way, of course. A thought experiment is different from a real-world viable machine.

In George Dyson's original posting, he said, "It is that primordial soup of archaic subroutines ... that is driving the push towards the sort of fault embracing template-based addressing that proved so successful in molecular biology".

If the question is framed in the future tense, then I understand what conversation we are having. (We're asking if evolving machines could hypothetically come to be in the future, perhaps the very far future.) I think this idea can be examined, and as I hope I made clear, I am open minded about it, although I maintain that an excessive emphasis on this possibility has negative effects on contemporary technology design and culture.

In more recent correspondence, George said quite plainly that, with regard to gaining autonomy through evolution, machines, "have done so *already*".

This I truly cannot accept. If people stopped maintaining today's machines they would not only cease to change, they would cease to operate entirely. I'm sure George must agree with that- that evolution based on small variations (mutations) allowed by error correction is not a possibility in machines as they exist today. So George must be

talking about a system made of people and computers together. And here, certainly, I think we must agree that there is room for alternate interpretations- that one person's autonomous machine might equally well be another's machine with an inscrutable user interface. If we can agree on this chain of reasoning, then I would hope to discuss whether there are pragmatic reasons to favor one interpretation over the other in specific circumstances, such as our own.

In correspondence, George suggested that we should start to think of the internet as already being somewhat autonomous, since it runs even though people don't fully understand it anymore. (I hope I'm doing justice in my paraphrasing.)

My experience of current digital tools is that while there are certainly numerous instances in which people no longer understand the tools, it is also true that these are precisely the same instances in which the tools fail- in which they crash. The changes that result from a human observing a crash are usually not incremental mutations, searching a space blindly for better configuration, but rather analysis-driven adjustments that force the machine to conform to a rational plan that was written down prior to testing. The plan might change, of course, but only on the human side of the system. I am not claiming that this is always the way that debugging happens (in fact I love to make little virtual worlds with quirky bugs I don't quite understand), but I am claiming that it is more true the larger a system gets.

The fact that Y2K bugs didn't destroy the world as feared is one piece of evidence that we are actually in charge of our machines, even though we like to fantasize that we aren't.

The examples I gave of people "making themselves stupid" in order to make software seem smart, as in the credit rating system, are ones in which people most definitely do understand the machines, to a fault.

The Internet as a machine seems comprehensible to me. At Advanced Network and Services, where the Internet 2 Engineering Office is located, and which is my primary perch these days, there's a fine project to measure activity on the net with probes all over the world, and the data are useful for rationally improving performance. No alien communication signals have appeared.

The failure modes of practical software are quite different from what is seen in chemical/biological systems. When a computer crashes (and I mean a real computer, not a thought experiment in a math journal), nothing else happens. The is no more processing. When an organism crashes, it turns into food for other organisms. Its information is not entirely lost from the system. I recognize that this point will probably fall on deaf ears to respondents who think of computers as already being autonomous and biological in some sense. I think a careful examination of computers as they are in the real world will show that all the "biological" properties of digital technology are brought to the table by the people who maintain the technology.

I don't think we know enough yet to say definitively whether the two kinds of unreliability (digital and biological/statistical) are ultimately, at some extreme of scaling, interchangeable.

I also don't perceive the evolution that George does in some of the examples he suggested in correspondence. In what ways have operating systems gotten better since the 70s? There are a few, but far fewer than anyone in the field ever imagined there would be. UNIX was, to a remarkable degree in retrospect, pretty much there at the start. I suppose it comes down to a subjective evaluation of how important various modifications since then have been.

The internet might provide better examples of the kinds of ongoing "evolution" George is talking about. There are still opportunities to create useful new subsystems, along the lines of the one operated by Akamai, for example. As another example, the TCP/IP protocol is probably the most common "soft failure mode" protocol in use, and it has improved over time, most notably with the advent of "slow start". But this happened when a human, Van Jacobson, had one of those thus far inscrutable "aha!" moments.

Ironically, I have for a long time nurtured a scheme to build an operating system out of components that would bind together using a pattern recognition approach (with so-called "neural nets") instead of literal reference, as part of my own war against "brittleness". Such a system, if I could ever get it to work, and I've tried, believe me, would be more in line with the Dysons' take on software than other architectures I am aware of out there in the real world today. (One sub-project of the Tele immersion Initiative, bearing the acronym SOFT, which has been created in the last two years at the Computer Science Department of Brown University, could perhaps be seen as an early example of a "soft binding" architecture.)

To Cliff Barney:

Hey, I'm thinking as socially as I can. Wish it were social enough for you!

I gave the closing talk at Stanford University's Englebart event that you mention. I presented a condensed version of the "missing half" of the manifesto there, and it's available on video (see http://unrev.stanford.edu/index.html). My preternaturally angelic and patient publishers are confident that I will somehow, someday soon finish the long overdue book that will unite both halfs.

Human society didn't change all THAT much during the course of the million-fold increase in computer power that you identify, from 1968 to roughly the end of the century. Certainly society changed more (as a result of technological provocation) in the previous 30 years, which saw the introductions of television, the birth control pill, factory-based genocide, the atomic bomb, LSD, the electric guitar, suburbia, the freeway, the middle class, and so much more. Globalism isn't all that new either. You can read passages in Marx on the internationalization of capital that sound exactly like dot com press releases from the recent boom years.

The last thirty years have seen such things as the rise of Gay rights and working moms, but it seems to me that many of these changes are most easily interpreted as extensions of processes that began before 1968. (As an example, I'm amazed that so much of today's teenage culture is as similar as it is to that of the 1950s and 1960s. The (white) music even sounds about the same as it did in the 1960s. The music of 1968 sounded quite different from the music of 1938.)

People talk about digital technology more than they use it. They tend to overstate how much they have been effected by it. I don't say this as a criticism. It's a most fascinating thing to talk about. Here I am doing it.

I think what's going on is that digital technology does not effect the lives of people until new culture, expressed both in software implementations and in changing human habits, is invented for it. Non-digital technologies, on the other hand, present instant opportunities for meaningful events to take place. Point a movie camera at the world and that world is changed forever, even if an initial subject is nothing more than an approaching train. Digital technology is different because an intensely time consuming process must precede its efficacy. An excessive degree of conscious forethought (thwarting pretensions to Dyonesian digital flights of fancy) and cumulative boredom characterize digital culture more than surprising revelation. The tedium gets to us all once in a while, and I think intellectual positions such as George Dyson's might serve as psychic comfort.

I am a true believer in the long term, lovely improvement of the human condition to be brought about by digital technology, but it's going to be a slow ride, because we have to build the code, piece by piece.

To Bruce Sterling:

A warm, brotherly bear hug for you!

To Rodney Brooks:

Your way of thinking is all too familiar, the standard issue point of view found in elite computer science departments. Glad you showed up, just in case anyone might have wondered if I was making up a straw man.

I made no claim as to whether machines could in theory become conscious or not. Instead I argued that such ultimate questions are not answerable, at least by anyone in our contemporary conversation.

I maintain, once again, that the most useful conversations we can have on such topics must be motivated by pragmatic, esthetic, and moral considerations.

Your certainty that you alone can identify the one true null hypothesis is a religious claim.

I hope it's clear that I was being snide and flip when I brought up nanobots. They are actors in a thought experiment, no more meaningful than artificial intelligence, and no more useful in thinking about how to design real machines, societies, and philosophies.

To Henry Warwick:

I'd like to address a plea to you and to other people who largely agree with me. Would you consider becoming immersed for a time in the other side's arguments, if only for the sake of dialog? They aren't stupid ideas, they're just wrong, and they deserve respect as smart, wrong ideas. If we humanists aren't willing to engage the CT crowd on their own terms once in a while, we can hardly expect them invest in understanding our terms.

I'd also suggest decoupling such questions as whether the universe is deeply "mathematical", or whether it can be fully understood, from the design, legal, esthetic, and social levels where the ideas that root in the heads of technologists come to matter.

The deep questions might never be answered. They must be asked, of course, but it is best to ask them separately. The pragmatic questions can not only be answered, but will be answered by our collective actions, whether we like it or not.

To Kevin Kelly:

I wrote the essay for my colleagues in the technology world, such as Rodney Brooks. Whether any of them are persuaded by it remains to be seen. My sense of this world is that it is currently not benefiting from a variagated ecology of metaphors, but rather is locked into a standard release of one metaphor.

To Margaret Wertheim:

I agree. Once Western culture defined itself as being on a ramp, the ramp had to go somewhere. The "other half" of the manifesto will be concerned with alternate ways of conceiving of the ramp's destination.

To John Baez:

Thank you for pointing out that a lot of folks in the "extropian" crowd seem to actually like the idea of goo taking over. I have come across this sentiment again and again. It is interesting in its own right, completely aside from whether Genghis Goo is a realistic scenario or not.

To Lee Smolin:

Thank you for this fascinating post.

I wish Stuart Kauffman would name his objects something other than "autonomous agents", since that is almost the same language CTers use to describe such things as the idiotic dancing paper clip that confuses users of Windows.

I'd like to encourage other respondents to address your ideas directly, instead of dragging the conversation down once again into eternal imponderables.

Some of the next deep (askable) questions: Will we someday be able to estimate how efficient natural evolution has been, in comparison to a theoretical ideal? Is evolution close to being as fast as it could be in searching the configuration spaces at hand, in the way that retinas are almost as sensitive to visible light as they could possibly be, or is there a lot of room for making evolutionary machines that would search practical configuration spaces much more quickly?

I'm also struck by how much more past computation is implied in some configurations than in others, and therefore wonder how your ontology relates to the various definitions of "information". Irreducible overhead in optimizing a configuration space (including legacy effects) might also be treated as a fundamental "distance" between configurations, and might serve as a basis for formal definitions of such things as species boundaries. This type of distance is also similar to some ideas about physical distance in recent computation quantum gravity models.

To Stewart Brand:

Yes, yes yes! This is the explanation for the preponderance of exceedingly strident expressions of libertarian ideals in digital culture.

To Daniel Dennett:

You'll be happy to know I turned down Harpers Magazine and instead accepted Wired's offer to print the .5 Manifesto. I assure you I am in no danger of drowning in a

friendly tsunami of Euro-admirers, for the simple reason that I am also a composer, and therefore the class of professional culture critics is sworn by blood oath to make my life difficult.

I'd like to be able to assert that neither of us understand something without being accused by CTers of sentimental, softheaded, retrograde religious dependency. I made no claim that there could never be an explanation for how people think, just that Darwin alone might not provide the framework for an explanation. No "half a skyhook", just an unsolved problem.

Straw men?

Read Rodney Brooks' posts and you'll see what I'm up against.

The rape book is silly, you just have to admit it. I could have quoted from dozens of clunkers in this odd text. There was a great passage about a woman raped by an orangutan who's husband (the woman's husband, that is) as well as she herself reported less consternation than they would have expected to experience if she had been raped by a person. No control group, sample size of one, reliance on subjective reportage, suspicious story; you could hardly come up with a more lousy experiment. And yet this example was used to reinforce the idea that the real reason rape is disliked is selfish genes; that bestiality is relatively delightful because it doesn't interrupt human mating schemes. I'm not saying, and have never said, that the ideas in this book are completely or exactly wrong, but rather that the book is inept. I sympathize with your position. You're a little like a member of a political party who has to defend an incompetent candidate. The important question to ask here is whether the CT community is too self-satisfied. I haven't met the authors of the rape book, but I imagine they must be intelligent and well meaning, and that perhaps the giddy team spirit of CT blinded them and made them sloppy.

I didn't attack Dawkins in the piece, and in fact a genial debate between he and I has been published. He is, as I have pointed out in past writings, not a meme totalist, even though he spawned a generation of them. As for you on consciousness, I am gently teasing you, and you must admit that you have been quite a rough player in your own writings in the past.

To Philip W. Anderson:

Thank you for your provocative note.

An interesting thought experiment is to imagine what the history of science and civilization might have been like if digital computers had become practical before Newton. This is not an unimaginable sequence. The ancient Alexandrians or Chinese might have done it if fortune had granted either of them a millennium or so of tranquility. The Chinese scenario might be more likely, since they weren't thinking in terms of mathematical proof, but were very good at coming up with clever technologies and building massive works. They would perhaps have built stylish city block-sized medieval computers out of electromechanical switches. These would have emitted marvelous rhythms, and perhaps there would have been dancing on the sidewalks around them.

I suspect our counterfactual predecessors could have gotten to the moon, but not built semi-conductors or an atomic bomb. They wouldn't have been forced to notice the problems that lead us to understand relativity and quantum mechanics.

I think there would have been less of a divide been the sciences and the mainstream of society, because it is easier to write fresh and fun computer programs than it is to do original work in continuous mathematics. Instead of being shrouded in esoteric mystery, science and engineering would have seemed more accessible to the lay person. Kant or his equivalent would have built huge simulations of competing metaphysics instead of seeking proofs.

Back to the present: Computers might yet yield important new physics. Stephen Hawking simply made the usual error of underestimating the time it takes to figure out how to write good software. We shouldn't expect deep understanding of software to improve any faster than deep understanding of other things. Think of the time it took to move from Newton to Einstein. Intellectual progress is not governed by Moore's Law.

Postscript:

Re: Ray Kurtzweil

Much to my surprise, Ray Kurtzweil and I spoke in succession (in Atlanta, at one of Vanguard's events) just as I was writing these responses. We see the world quite differently. He would certainly reject my last claim above, that fundamental intellectual achievement isn't inexorably speeding up.

I see punctuated equilibria in the history of science. Right now we're in the midst of an explosion of new biology. Around the turn of the last century there was an explosion of data and insight about physics. Physics is now searching for its next explosion but hasn't found it yet.

I also see a distinction between quantity and quality that Ray doesn't. I see computers getting bigger and faster, but it doesn't directly follow that computer science is also improving exponentially.

Ray sees everything as speeding up, including the speed of the speedup. In Atlanta, he collected varied graphic portrayals of exponential historical processes in a slide show, and labeled these a "countdown" to the singularity he predicts will arrive about a quarter of the way into the new century.

His exponential histories blend what others might think of as varied phenomena together into categories without differentiation. For instance, he showed a slide about Moore's Law, but with the timeframe not limited to the era of the silicon chip. Instead, he defines chips as just one of five technological phases that have upheld the exponential speedup of computation that started with the earliest mechanical calculation devices. He infers that the curve will be continued with nanotechnological or other devices once the limits of chip technology are reached, in perhaps twelve years. Likewise he showed a grand exponential account of the history of life on Earth that started with items like the Cambrian Explosion at the foot of the curve and soared to modern technological marvels at its heights, as if these were all of a kind.

I hope I can avoid being cast as the person who precisely disagrees with Ray, since I think we agree on many things. There are exponential phenomena at work, of course, but I feel they have robust contrarian company. I believe our human story is not best defined by a smooth curve, even at a large scale (although I try to make one exception, which I'll describe below). If there was ever a complex, chaotic phenomenon, we are it.

One question I have about Ray's exponential theory of history is whether he is stacking the deck by choosing points that fit the curves he wants to find. A technological pessimist could demonstrate a slow-down in space exploration, for instance, by starting with sputnik, and then proceeding to the Apollo and the space shuttle programs and then to the recent bad luck with Mars missions. Projecting this curve into the future could serve as a basis for arguing that space exploration will inexorably wind down. I've actually heard such reasoning put forward by antagonists of NASA's budget. I don't think it's a meaningful extrapolation, but it's essentially similar to Ray's arguments for technological hyper-optimism.

It's also possible that evolutionary processes might display local exponential features at only some scales. Evolution might be a grand scale "configuration space search" that periodically exhibits exponential growth as it finds an insulated cul-de-sac of the space that can be quickly explored. These are regions of the configuration space where the vanguard of evolutionary mutation experimentation comes upon a limited theater within which it can play out exponential games like arms races and population explosions. I suspect you can always find exponential sub processes in the history of evolution, but they don't give form to the biggest picture.

Here's one example: The dinosaurs were apparently "scaled" (maybe in both the traditional and Silicon Valley senses of the word!) by an "arms race", leading to larger and larger animals. Dinosaurs were not the only creatures at the time that relied on gigantism as a strategy. Much of the animal kingdom was becoming huger at once. I doubt the size competition proceeded at a linear rate. Arms races rarely do.

If we were dinosaurs debating this question, the Kurtzweilosaurus might argue that our descendants would soon be big enough to stand on their toes and touch the moon, and not long after that become as big as the universe. (Tribute is due, as always, to Mark Twain and his erectile Mississippi.)

The race to bigness came to a halt, perhaps because of a spaceborne cataclysm. Whatever the reason for the dinosaurs' disappearance, they could not have become bigger without bounds. Furthermore, the race to bigness did not inexorably reappear, but was replaced by other races. The mere appearance of an exponential sequence does not mean that it will not encounter an impassable boundary, or become untraceable as other processes exert their influences.

I see a scattered distribution of local, bounded exponential processes in the history of life, while Ray sees these processes all focusing like a coherent laser on a point in time we will likely live to see.

Smart people can be fooled by trends. For instance, in 1666, when technological optimism was perhaps even more pronounced than it is today (when space exploration

seemed to be progressing exponentially, for instance), Time Magazine presented what it thought was a sober prediction: That by the year 2000 technology would have advanced to the point that no one in America would work for a living. Automation would take the drudgery out of life. Each American citizen would receive a healthy middle class stipend in the mail every month simply for being American. A specific dollar amount ($30-$40,000 in 1966 dollars) was even projected for the stipend. (Thanks to GBN's Eamonn Kelly for pointing out this example.)

Time Magazine was making what it saw as a perfectly reasonable extrapolation based on legitimate data. What went wrong with Time's prediction? There's no doubt that technology continued to improve in the second half of the twentieth century, and by most interpretations it did so at an exponential clip. Productivity faithfully increased on an exponential curve as well.

Here are a few candidate failings: Public rejection of key predicted technologies such as nuclear energy; "lock in" of such things as cars and freeways, which did not scale cheaply or elegantly; population explosions; increasingly unequal distributions of wealth; entrenchment in law and habit of the work ethic; and perhaps even the beginning of the "planet of helpdesks" scenario that made a cameo appearance in the .5 manifesto. This last possibility provides an alternate way to think about the growing "knowledge economy".

Note that some of these countervailing elements are exponential in their own right. Population growth is a classic example of an exponential process that can absorb an exponential increase in available resources. This is what has happened with high yield agriculture in India.

What's really tricky is figuring out when one process will outrun its surroundings for a while in a meaningful way, as the Internet has grown at a faster rate than the population or the larger economy.

I have to admit that I want to believe in one particular large scale, smooth, ascending curve as a governor of mankind's history. Specifically, I want to believe that moral progress has been real, and continues today. This is not an easy thing to believe in. I formed my desire to believe in it at about the same that Time Magazine made it's prediction about the end of work.

I remember being a child in the 1960s, and there was a giddy feeling in the air of accelerating social change. While the language was different, the idea wasn't that different from today's digital eschatology. It felt like the world was on an exponential course of change, approaching a singularity.

The evidence was there. You could have plotted the points on a graph and seen one of Ray's curves, but no one thought to do it explicitly at the time. 1776, Civil War, Women's Suffrage, Civil Rights Struggle, Anti-war movement, Women's lib, Gay Rights, Animal rightsSÂ You could plot all these on a graph and see an exponential rate of expansion of the "Circle of Empathy" I wrote about in the .5 Manifesto. This process seemed to be destined to zoom into a singularity around 1969 or so, when I was nine years old. People were quite depressed when the singularity did not happen. Younger

people today might not realize how deeply that singularity's no-show marked the lives of a vast number of Baby Boomers.

Dinosaurs did not become as large as the universe, work did not disappear in 2000 (at least not by November, 2000, as I write this), and love did not conquer all in 1969. All the trends were real, but were either interrupted, outran their own internal logics, ran out of world to expand into, or were balanced or consumed by other processes.

Back to "One Half of a Manifesto by Jaron Lanier; Reality Club Comments

Re: Ray Kurtzweil

Much to my surprise, Ray Kurtzweil and I spoke in succession (in Atlanta, at one of Vanguard's events) just as I was writing these responses. We see the world quite differently. He would certainly reject my last claim above, that fundamental intellectual achievement isn't inexorably speeding up.

I see punctuated equilibria in the history of science. Right now we're in the midst of an explosion of new biology. Around the turn of the last century there was an explosion of data and insight about physics. Physics is now searching for its next explosion but hasn't found it yet.

I also see a distinction between quantity and quality that Ray doesn't. I see computers getting bigger and faster, but it doesn't directly follow that computer science is also improving exponentially.

Ray sees everything as speeding up, including the speed of the speedup. In Atlanta, he collected varied graphic portrayals of exponential historical processes in a slide show, and labeled these a "countdown" to the singularity he predicts will arrive about a quarter of the way into the new century.

His exponential histories blend what others might think of as varied phenomena together into categories without differentiation. For instance, he showed a slide about Moore's Law, but with the timeframe not limited to the era of the silicon chip. Instead, he defines chips as just one of five technological phases that have upheld the exponential speedup of computation that started with the earliest mechanical calculation devices. He infers that the curve will be continued with nanotechnological or other devices once the limits of chip technology are reached, in perhaps twelve years. Likewise he showed a grand exponential account of the history of life on Earth that started with items like the Cambrian Explosion at the foot of the curve and soared to modern technological marvels at its heights, as if these were all of a kind.

I hope I can avoid being cast as the person who precisely disagrees with Ray, since I think we agree on many things. There ARE exponential phenomena at work, of course, but I feel they have robust contrarian company. I believe our human story is not best defined by a smooth curve, even at a large scale (although I try to make one exception, which I'll describe below). If there was ever a complex, chaotic phenomenon, we are it.

One question I have about Ray's exponential theory of history is whether he is stacking the deck by choosing points that fit the curves he wants to find. A technological

pessimist could demonstrate a slow-down in space exploration, for instance, by starting with sputnik, and then proceeding to the Apollo and the space shuttle programs and then to the recent bad luck with Mars missions. Projecting this curve into the future could serve as a basis for arguing that space exploration will inexorably wind down. I've actually heard such reasoning put forward by antagonists of NASA's budget. I don't think it's a meaningful extrapolation, but it's essentially similar to Ray's arguments for technological hyper-optimism.

It's also possible that evolutionary processes might display local exponential features at only some scales. Evolution might be a grand scale "configuration space search" that periodically exhibits exponential growth as it finds an insulated cul-de-sac of the space that can be quickly explored. These are regions of the configuration space where the vanguard of evolutionary mutation experimentation comes upon a limited theater within which it can play out exponential games like arms races and population explosions. I suspect you can always find exponential sub processes in the history of evolution, but they don't give form to the biggest picture.

Here's one example: The dinosaurs were apparently "scaled" (maybe in both the traditional and Silicon Valley senses of the word!) by an "arms race", leading to larger and larger animals. Dinosaurs were not the only creatures at the time that relied on gigantism as a strategy. Much of the animal kingdom was becoming huger at once. I doubt the size competition proceeded at a linear rate. Arms races rarely do.

If we were dinosaurs debating this question, the Kurtzweilosaurus might argue that our descendants would soon be big enough to stand on their toes and touch the moon, and not long after that become as big as the universe. (Tribute is due, as always, to Mark Twain and his erectile Mississippi.)

The race to bigness came to a halt, perhaps because of a spaceborne cataclysm. Whatever the reason for the dinosaurs' disappearance, they could not have become bigger without bounds. Furthermore, the race to bigness did not inexorably reappear, but was replaced by other races. The mere appearance of an exponential sequence does not mean that it will not encounter an impassable boundary, or become untraceable as other processes exert their influences.

I see a scattered distribution of local, bounded exponential processes in the history of life, while Ray sees these processes all focusing like a coherent laser on a point in time we will likely live to see.

Smart people can be fooled by trends. For instance, in 1666, when technological optimism was perhaps even more pronounced than it is today (when space exploration seemed to be progressing exponentially, for instance), Time Magazine presented what it thought was a sober prediction: That by the year 2000 technology would have advanced to the point that no one in America would work for a living. Automation would take the drudgery out of life. Each American citizen would receive a healthy middle class stipend in the mail every month simply for being American. A specific dollar amount ($30-$40,000 in 1966 dollars) was even projected for the stipend. (Thanks to GBN's Eamonn Kelly for pointing out this example.)

Time Magazine was making what it saw as a perfectly reasonable extrapolation based on legitimate data. What went wrong with Time's prediction? There's no doubt that technology continued to improve in the second half of the twentieth century, and by most interpretations it did so at an exponential clip. Productivity faithfully increased on an exponential curve as well.

Here are a few candidate failings: Public rejection of key predicted technologies such as nuclear energy; "lock in" of such things as cars and freeways, which did not scale cheaply or elegantly; population explosions; increasingly unequal distributions of wealth; entrenchment in law and habit of the work ethic; and perhaps even the beginning of the "planet of helpdesks" scenario that made a cameo appearance in the .5 manifesto. This last possibility provides an alternate way to think about the growing "knowledge economy".

Note that some of these countervailing elements are exponential in their own right. Population growth is a classic example of an exponential process that can absorb an exponential increase in available resources. This is what has happened with high yield agriculture in India.

What's really tricky is figuring out when one process will outrun its surroundings for a while in a meaningful way, as the Internet has grown at a faster rate than the population or the larger economy.

I have to admit that I want to believe in one particular large scale, smooth, ascending curve as a governor of mankind's history. Specifically, I want to believe that moral progress has been real, and continues today. This is not an easy thing to believe in. I formed my desire to believe in it at about the same that Time Magazine made it's prediction about the end of work.

I remember being a child in the 1960s, and there was a giddy feeling in the air of accelerating social change. While the language was different, the idea wasn't that different from today's digital eschatology. It felt like the world was on an exponential course of change, approaching a singularity.

The evidence was there. You could have plotted the points on a graph and seen one of Ray's curves, but no one thought to do it explicitly at the time. 1776, Civil War, Women's Suffrage, Civil Rights Struggle, Anti-war movement, Women's lib, Gay Rights, Animal rightsŠ You could plot all these on a graph and see an exponential rate of expansion of the "Circle of Empathy" I wrote about in the .5 Manifesto. This process seemed to be destined to zoom into a singularity around 1969 or so, when I was nine years old. People were quite depressed when the singularity did not happen. Younger people today might not realize how deeply that singularity's no-show marked the lives of a vast number of Baby Boomers.

Dinosaurs did not become as large as the universe, work did not disappear in 2000 (at least not by November, 2000, as I write this), and love did not conquer all in 1969. All the trends were real, but were either interrupted, outran their own internal logics, ran out of world to expand into, or were balanced or consumed by other processes.

Henry Warwick

Artist, composer, and scientist

Responding to all of Mr Lanier's lengthy Manifesto would make for an enormous essay several times longer than his Manifesto. Rather than engage in lengthy interplay of point by point analysis, my contribution to this discussion will first set out what I believe/perceive to be true, then go into my own prognosis of the future, specifically the anti — utopian vision of what Mr Lanier calls Cybernetic Totalism. I call it delusional technocratic arrogance, but I won't quibble about that. In deference to his essay, I'll refer to it as "CT"...

Mr Lanier sets out (what he believes) are six component beliefs of CT. I think it's actually much simpler than that, and it fundamentally breaks down into a basic core group of related beliefs/predictions:

Someday, soon, we will either replace ourselves or be replaced by robots/computers.

Failing (or in addition to) that, we will divide the human genome into an enhanced variety and the rest ["archaic"] of humanity.

Related to #2, we might also divide the race off as bio — mechanical creatures, what I call the "Borg Fantasy"

Point 1 will never happen — because it can't.

Point 2 will happen, but the results will probably be different than we envision, and the timing on it will likely be much later than sooner.

Point 3 won't happen, as the extreme variety as envisioned by various contemporary fantasies like Star Trek's Borg are just plain stupid, and while future technologies will help us in many ways, especially in terms of communications, incorporating them as body parts seems inherently dimwitted given Moore's Law. Attaching or putting machines into ourselves just doesn't make a whole lot of sense.

The rest of Mr Lanier's discussion is spent blasting their theoretic superstructure. As admirable such an effort may be, I see it as unnecessary, much as it is unnecessary for a democrat to argue Courtly Manners with a monarchist. The point is the sham of the divine right of kings, not whether bowing is bad for your back.

So, directly to a basic point — beneath the CT position is a fundamental and unspoken axiom — the Pythagorean Conjecture that the universe is mathematical, and deeper still, that the universe is fundamentally understandable by humans. Pythagoras took it to a numerological extreme, but the fundamental myth still obtains with many people who work in science — everyone is looking for the Equation/Theory/axiomatic system that will explain Everything Forever. The CT position depends on this assumption. Yet, we have never had, nor do we have now, any conclusive proof that the universe is humanly understandable in the first place, much less representable in some reductivist symbology of mathematics or any other language for that matter. Indeed, with Godel et al, we have a number of theories demonstrating the very limitations of such endeavors in the first place.

The CT position assumes that the world is computable and their thinking machine project logically follows — logical machines for a logical universe.

My thinking is this: The Universe is beyond human comprehension,

[Re: Haldane: “The Universe is not only weirder than you think — it’s weirder than you can think” and Brockman: “Nobody knows and you can’t find out.”]

and is therefore not computable.

However — because of our inquisitive nature and history of inquiry and Inquisition, we have to continue the effort of the Scientific Project — just because there is no possibility of coming to a complete understanding and total knowledge of everything doesn’t mean that we can’t come to understandings that are useful and provide a reasonable and coherent sense of the universe and its workings, given our limited capacities to understand it. For example — cultural anthropology is a program that can never be finished, because there are cultures that would be changed irrevocably or destroyed just upon their being observed by the Western Cultural Anthropology Industry. Does this mean that anthropologists should pack up their tents and surrender? Of course not. The same goes for all the other disciplines of science. The Scientific Method works extremely well — we should keep that — but we should be more humble in matters regarding our actual abilities, as we use the Scientific Method to expand such abilities.

Once we abandon this obsessive fanaticism of absolute complete knowledge, we can continue on with our process of discovery without the headache of a deadline. Knowing there are actual limits allows us to push our perceived limits in a way that we can pick and choose our battles with the Great Unknown. When we give up on knowing every detail of the Mind of some imaginary Friend, we might acquire some of our Friend’s imagination and wisdom.

Now, regarding the “We Will Have A Sentient Machine by 2030 and We Will All Be Replaced By Robots or Evolve Into The BORG” nonsense —

First off, the notion is so blatantly Millennialist and stupendously lacking in imagination, I find it rather sad that otherwise intelligent and reasonable people actually hold such paranoid malarkey as a position worth defending with as much vigor as they do. I dismiss the CT prediction wholesale directly for that general reason. They remind me of an old encyclopedia I found in the trash as a young boy. It was from 1927, and when I found it, it was 40 years old and looked 100. It contained some illustrations of what a city in the year 2000 would look like — giant skyscrapers separated by 20 lane highways, dozens of dirigibles and hundreds of airplanes flying between the buildings. Factories were invisible, and there wasn’t a toxic waste dump in sight. I see the CT predictions in much the same way. Yes, air travel expanded a lot since 1927, and while we don’t have many dirigibles, the NJ Turnpike and the 5 in LA certainly qualify as giant highways. The same will go for AI in 2030. To the chagrin of the CTs, machines won’t think (because they can’t) but thanks to the tireless efforts of the CTs, computers will do a lot of useful work for us, even more so than now, and will invent entire new categories of productive labor for humans.

In general, I find the CT position laughable and tragic. In specific, there are other points regarding their philosophic superstructure they have erected to defend their position that should be addressed.

First the Turing Test.

My objection to the Turing Test is this:

The very basis of the Turing test is one that knows that machines don't think — the whole thing is based on deception.

Who is to do the judging?

As far as the judges go — I'm sure as hell not going to trust the geeks who make the first "thinking machine" to tell me it's really thinking. I may be eccentric and a bit deranged, but I'm not that stupid. Yet.

Regarding point 1 — Deception:

There's one thing the CT crowd really doesn't want to accept — that they are deceiving themselves, and the Turing Test is the tool of their deception.

Fact: Machines don't and can't think. Existence precedes essence. Computers pass voltages. Period. They don't remember anything. They don't think about anything. Everything we discuss or sense about them is secondary and something we bring to it. Saying that computers think is like discussing the political persuasions of rock formations.

Once we see the Turing Test for what it really is, the real CT/Turing Test project is now revealed:

Can we make machines operate in such a way that we can — deceive — ourselves into thinking there's actually a sentient human in it? And can we deceive/bludgeon others into agreeing with us?

The Turing Testers know that machines aren't sentient, as they wait for the next rev of some machine to trick them. And once "tricked" — what makes them think they or anyone else wouldn't know it's a trick — everytime?

"Gee — last week, the HAL 9000 passed the Turing test. Well wuddya know — that last algorithm really did the trick. Let's check it out now. UhOh. Today it's not passing the Turing test…so I guess it isn't sentient anymore…"

That we so deceive ourselves does not mean the condition of sentience is or ever was actually present — it simply means that the required conditions to our test have been met at a particular historical juncture — on a given day, the machine has "fooled" us into thinking it can think. It's been programmed in such a way that we are led to believe it has a mind. This doesn't mean it actually has one. With the Turing Test, the machine must simply be able to do what we expect of a human within a certain range of activity. But is it Sentient? Hell No. It doesn't take Albert Einstein to see how nekkid that Emperor is.

Another objection I have to the Turing Program is — why bother? Humans are such a contemptible lot of petty, ignorant, messy, obscene and violent whiners, why would we ever want to make a computer act like one? I find the idea of simulating human behavior so ludicrous, it's appalling that it has occupied so much airtime for the past several decades. It's a sad testament to our ignorance and vanity.

However, this doesn't mean that machines that attempt to simulate awareness can't do useful work. On the contrary, I am firmly convinced that they can and should do the work that humans are simply not designed for — space exploration, deep sea work, and a thousand other extremely dangerous but mission critical activities.

Lanier is right on the money with his circle of empathy. The computer might whine, complain, threaten violence, whatever. Just unplug the thing. It's not a person, it's

not sentient, and people who think it is need some guidance on personal boundary conditions.

So, the Robot/Computer Mind isn't going to happen, because it can't. We will have machines that can do some amazing things, but sentience ain't one of 'em.

Now, for the other points — the biological and the biomechanical.

Assuming homo sapiens doesn't go extinct without issue, *homo futurus* is inevitable. It's not a question of if; it's a matter of when and how. If civilization goes completely belly up, and we're all reduced to wandering bands of nomads in a world filled with toxic waste dumps and highly oxidized metal particles, *homo futurus* will be a hardy and tough human built for life on the run hunting the giant *rattus futurus* for food and avoiding the roving psychotic packs of *rotweilerus futurus*. It'll be a tough life, but not without its rewards. We will evolve to adapt to such circumstances.

If we get some collective sense in our skulls, and reduce our numbers to a sustainable value (200 — 400 million?) with our science we can eventually biologically enhance ourselves into a *homo futurus* — a creature of our own design. Socially speaking, the introduction of such technology would be simple enough — if someone said,

"Mr and Mrs Warwick — the next girl you have will live to be about 140 and die with the body of a 50 year old, have 20/10 eyesight including some sensitivity in the infrared and UV spectra, have hearing between 5 Hz and 60kHz and she'd look like a buff cross between Marilyn Monroe and Katherine Hepburn who never sunburns, and be able to dance better than Ginger Rogers and have an IQ in the high 4 digits all for only $260,000 please sign at the bottom in ink please."

We'd be there signing paper with my Pelikan in a New York Second. We'd also be in hock for the rest of our comparatively short lives, but we'd do it in a heartbeat, and I think many other people would too.

So, the biological working of the species will, I believe, be inevitable as we learn more and more about the human genome. However, I don't think this level of understanding will be any time soon. If we're diligent and work hard — maybe we'll have it in a few hundred years. I imagine there will be a bunch of people opposed to it on "ethical" grounds, and I can't imagine what the test trials would be like, but eventually it will happen if DNA technology keeps a pace even vaguely resembling Moore's Law.

I'm not too concerned about having two species around, either — as these treatments become more commonplace, they'll go down in price, and if different companies compete, we could have the most enhancement for the lowest price, and most every genetic line/family will eventually be able to have their progeny continue into the next phase of human evolution. In fact the later adopters might even have some advantages compared to the early adopters. Like IQ in the high FIVE figure range…cool…

As I said — barring extinction of sapiens, *homo futurus* is inevitable. It's not a matter of if; it's just a question of when and how, and it could be a Very Good Thing — not a future to fear. There's also a non-zero chance *homo futurus* will be wearing

deerskin and chasing bunnies for dinner, but that's not something under our immediate control.

As far as the biomechanical future goes, I think that is a dubious future, as I the "Borg model" is absurd, paranoid, and juvenile. It's an irrational fantasy based more in Cold War politics than honest and reasonable technological conjecture. There's been a lot of discussion about nanotechnology and technological implants since Drexler et al back in the 1980s, but so far it's been mostly just that — a lot of discussion with only a few scattered developments, including a fellow with my last name in the UK. I don't think the research is bad — I just don't have any faith in its applications.

On the other hand, I do believe biomechanical devices could be of some use, especially if they are wearable — with technologies in the near future, email and voice communications could be as simple as wearing two small transducers that stick to the bony points behind your ears. Implants? They're so messy and atavistic.

And Finally —

There will be no Cybernetic Cataclysm in 2030, just like there was no Armageddon in 1999. Short of a wayward asteroid coming to visit and ruin an afternoon for a few million years, things generally don't work that way around here. We are far too good and earnest to deserve some techno Hell, and we're way too selfish and myopic to understand Heaven. The machines, as cheery and responsive as we might make them, aren't and won't be sentient. So, we're stuck here, alone but for our chimp cousins, on this little green planet. It's a nice place. We need to take care of it a lot better than we have been. We need to clean it up and invent a fun, clean, future. Send the machines into space — they can tell us of other planets. Maybe a few of us will go check out the nicer ones. Maybe even bring a few of our chimp cousins.

Don't worry about the Borg or the Forbin Project taking over. That only happens in lame Hollywood movies written for 15 — year — old boys. Frankly, I'm a lot more concerned about the very human Supreme Court repealing the Bill of Rights on the altar of the Permanent Wartime Economy, and how we're going to come up with the energy needed to run our machines, heat our homes, and cook our food, when the oil runs out.

Kevin Kelly

Senior Maverick, Wired; Author, What Technology Wants and The Inevitable

Jaron doesn't have to worry about the cybernetic metaphor, because he says his main concern is that it has become sole metaphor of our time, or at least the sole metaphor of our tribe. If that were really true, I'd worry too. But it isn't.

What the cybernetic metaphor is an extreme perspective, an inverted perspective that will eventually play out its usefulness. It is similar (and related) to Richard Dawkins' famous view of the selfish gene. Dawkins says that you can understand a lot which is new, and a re-understand a lot of the old orthodoxy, by looking at the world from the view of genes. In fact you can begin to look at everything that way, and for a while wherever you look, the world looks different. This view can unleash new understandings. What is important to remember is that while Dawkins looks at the world that way, this is not the only way he looks at it. In his daily life he adopts a quite ordinary view of the world. I have looked at the world in Dawkins selfish-gene way, and then the next minute

I have looked at the world in Jaron's way. Most of the time (but not all!) I see more new things via Dawkins way. I might also look at the world via Freud's way, or Marx's way, but I usually don't see much interesting to me that way.

The new cybernetic metaphor, on the other hand, is very powerful. We can look at almost anything now, from physics, to emotions, to nature, to experience itself, and find new things when we imagine it as computation. We can imagine people as robots and learn all kinds of things. I can do that one minute and then the next minute I can play with my little 4-year old boy, and see him only through the eyes of a naked primate. Eventually we (as a culture) will finish examining everything via the cybernetic metaphor, and then we'll get bored. But the important thing is that right now almost anything we examine will yield up new insights by imaging it as computer code. And -- this is important -- while one re-examines the world in this way, it is vital that you take the metaphor seriously. It should be the only metaphor you see while you are looking through it. The next minute we can adopt another view.

I think we have not come close to exhausting this metaphor, and as my earlier essay on it (called the Computer Metaphor) suggests, I think it will overturn our current ideas of physics and culture first, before we abandon it. It is dangerous, but not because it is our only tool.