

ONE-HALF OF A MANIFESTO

SCREED GRAB

This article has been reproduced in a new format and may be missing content or contain faulty links. Contact wiredlabs@wired.com to report an issue.

Why stupid software will save the future from neo-Darwinian machines.

For the last 20 years, I have found myself on the inside of a revolution, but on the outside of its resplendent dogma. Now that the revolution has not only hit the mainstream, but bludgeoned it into submission by taking over the economy, it's probably time for me to cry out my dissent more loudly than I have before.

And so I shared the following thoughts with the members of Edge.org, many of whom are, as much as anyone, responsible for this revolution, one which champions the ascent of cybernetic technology as culture. That first "One-Half of a Manifesto," as technology manifestos sometimes do, quickly blossomed across other Web sites and beyond.

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.

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 Wiener, 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. Wiener 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. Cybernetic patterns of information provide the ultimate and best way to understand reality.
2. People are no more than cybernetic patterns.
3. Subjective experience either doesn't exist, or is unimportant because it is some sort of ambient or peripheral effect.
4. What Darwin described in biology, or something like it, is in fact also the singular, superior description of all creativity and culture.
5. Qualitative as well as quantitative aspects of information systems will be inexorably accelerated by Moore's law.

And finally, the most dramatic:

6. 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 20 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 *Gödel, Escher, Bach* and growing more harsh with recent entries such as *The Age of Spiritual Machines* by Ray Kurzweil.

Recently, public attention has finally been drawn to number six, 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 imminent 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 nanobots might take over the planet in 20 years. It seems beneath their dignity. And yet, the eschatologies of Kurzweil, 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 part away from 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 number one

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 1970s. 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 endgame, 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 versus 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 number one appeared on the stage almost immediately with the first computers. It was articulated by the first generation of computer scientists: Wiener, 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 number two

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 remarketed 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 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 humanlike.

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. 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 important, 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 number three

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 that appear to cybernetic totalists to populate reality today, along with human beings. It is certainly true that we are confronted with nonhuman and metahuman 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-

or 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, 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 led 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 20th 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 forced 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, 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 or 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 extrarational 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 number four

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 posthuman 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 mean to imply that there is a completely unified bloc 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 toward 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: The Spread of Excellence From Plato to Darwin* that if there's an arrow in evolution, it's toward 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, but it is more important to point out an epistemological conundrum that should be considered by Darwinian eschatologists. 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 superbeings 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 cyberspecies 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 nonhuman 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 that 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: Biological Bases of Sexual Coercion* 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 Wilson 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 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 in 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 enormous 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 but seem to have sunk to an almost postmodern 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 describe in the next section.

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

Belief number five

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 '70s – 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 its 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 databases 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 unwieldy, 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 rely on an exponential extrahuman 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 nontechnical 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 key-press, than they did 15 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 data sets 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 gridlock 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 nonhuman 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 unwieldy only because of 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 Kurzweil, 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 unwieldiness 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 resulting 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 its 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 number six

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. (See "Why the Future Doesn't Need Us," *Wired* 8.04, page 238.) Joy accepts the pronouncements of Ray Kurzweil and others, who believe that Moore's law will lead to autonomous machines, perhaps by the year 2020. That is 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 quite a bag of tricks.

They'll be able to perform nanomanufacturing, 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 Joy 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 Joy'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 one 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 20 or 30 years, there will not only be a vast amount of computation going on planet Earth, but the maintenance of that computation will consume the efforts of almost every living person. We're talking about a planet of help desks.

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 dotcom 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 ultrarich 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 20 or 30 years they might become quite different indeed. Will the ultrarich 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 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 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 were to become cheap, so would the horrific biological weapons of Bill Joy'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 that 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.

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 videogames 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 not to 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.