CHAPTER 2

an apocalypse of self-abdication

THE IDEAS THAT I hope will not be locked in rest on a philosophical foundation that I sometimes call cybernetic totalism. It applies metaphors from certain strains of computer science to people and the rest of reality. Pragmatic objections to this philosophy are presented.

What Do You Do When the Techies Are Crazier Than the Luddites?

The Singularity is an apocalyptic idea originally proposed by John von Neumann, one of the inventors of digital computation, and elucidated by figures such as Vernor Vinge and Ray Kurzweil.

There are many versions of the fantasy of the Singularity. Here's the one Marvin Minsky used to tell over the dinner table in the early 1980s: One day soon, maybe twenty or thirty years into the twenty-first century, computers and robots will be able to construct copies of themselves, and these copies will be a little better than the originals because of intelligent software. The second generation of robots will then make a third, but it will take less time, because of the improvements over the first generation.

The process will repeat. Successive generations will be ever smarter and will appear ever faster. People might think they're in control, until one fine day the rate of robot improvement ramps up so quickly that superintelligent robots will suddenly rule the Earth.

In some versions of the story, the robots are imagined to be microscopic, forming a "gray goo" that eats the Earth; or else the internet itself comes alive and rallies all the net-connected machines into an army to control the affairs of the planet. Humans might then enjoy immortality within virtual reality, because the global brain would be so huge that it would be absolutely easy—a no-brainer, if you will—for it to host all our consciousnesses for eternity.

The coming Singularity is a popular belief in the society of technologists. Singularity books are as common in a computer science department as Rapture images are in an evangelical bookstore.

(Just in case you are not familiar with the Rapture, it is a colorful belief in American evangelical culture about the Christian apocalypse. When I was growing up in rural New Mexico, Rapture paintings would often be found in places like gas stations or hardware stores. They would usually include cars crashing into each other because the virtuous drivers had suddenly disappeared, having been called to heaven just before the onset of hell on Earth. The immensely popular *Left Behind* novels also describe this scenario.)

There might be some truth to the ideas associated with the Singularity at the very largest scale of reality. It might be true that on some vast cosmic basis, higher and higher forms of consciousness inevitably arise, until the whole universe becomes a brain, or something along those lines. Even at much smaller scales of millions or even thousands of years, it is more exciting to imagine humanity evolving into a more wonderful state than we can presently articulate. The only alternatives would be extinction or stodgy stasis, which would be a little disappointing and sad, so let us hope for transcendence of the human condition, as we now understand it.

The difference between sanity and fanaticism is found in how well the believer can avoid confusing consequential differences in timing. If you believe the Rapture is imminent, fixing the problems of this life might not be your greatest priority. You might even be eager to embrace wars and tolerate poverty and disease in others to bring about the conditions that could prod the Rapture into being. In the same way, if you believe the Singularity is coming soon, you might cease to design technology to serve humans, and prepare instead for the grand events it will bring.

But in either case, the rest of us would never know if you had been right. Technology working well to improve the human condition is detectable, and you can see that possibility portrayed in optimistic science fiction like *Star Trek*.

The Singularity, however, would involve people dying in the flesh and being uploaded into a computer and remaining conscious, or people simply being annihilated in an imperceptible instant before a new superconsciousness takes over the Earth. The Rapture and the Singularity share one thing in common: they can never be verified by the living.

You Need Culture to Even Perceive Information Technology

Ever more extreme claims are routinely promoted in the new digital climate. Bits are presented as if they were alive, while humans are transient fragments. Real people must have left all those anonymous comments on blogs and video clips, but who knows where they are now, or if they are dead? The digital hive is growing at the expense of individuality.

Kevin Kelly says that we don't need authors anymore, that all the ideas of the world, all the fragments that used to be assembled into coherent books by identifiable authors, can be combined into one single, global book. *Wired* editor Chris Anderson proposes that science should no longer seek theories that scientists can understand, because the digital cloud will understand them better anyway.*

Antihuman rhetoric is fascinating in the same way that self-destruction is fascinating: it offends us, but we cannot look away.

The antihuman approach to computation is one of the most baseless ideas in human history. A computer isn't even there unless a person experiences it. There will be a warm mass of patterned silicon with electricity coursing through it, but the bits don't mean anything without a cultured person to interpret them.

This is not solipsism. You can believe that your mind makes up the world, but a bullet will still kill you. A virtual bullet, however, doesn't

*Chris Anderson, "The End of Theory," *Wired*, June 23, 2008 (www.wired.com/science/discoveries/magazine/16-07/pb_theory).

even exist unless there is a person to recognize it as a representation of a bullet. Guns are real in a way that computers are not.

Making People Obsolete So That Computers Seem More Advanced

Many of today's Silicon Valley intellectuals seem to have embraced what used to be speculations as certainties, without the spirit of unbounded curiosity that originally gave rise to them. Ideas that were once tucked away in the obscure world of artificial intelligence labs have gone mainstream in tech culture. The first tenet of this new culture is that all of reality, including humans, is one big information system. That doesn't mean we are condemned to a meaningless existence. Instead there is a new kind of manifest destiny that provides us with a mission to accomplish. The meaning of life, in this view, is making the digital system we call reality function at ever-higher "levels of description."

People pretend to know what "levels of description" means, but I doubt anyone really does. A web page is thought to represent a higher level of description than a single letter, while a brain is a higher level than a web page. An increasingly common extension of this notion is that the net as a whole is or soon will be a higher level than a brain.

There's nothing special about the place of humans in this scheme. Computers will soon get so big and fast and the net so rich with information that people will be obsolete, either left behind like the characters in Rapture novels or subsumed into some cyber-superhuman something.

Silicon Valley culture has taken to enshrining this vague idea and spreading it in the way that only technologists can. Since implementation speaks louder than words, ideas can be spread in the designs of software. If you believe the distinction between the roles of people and computers is starting to dissolve, you might express that—as some friends of mine at Microsoft once did—by designing features for a word processor that are supposed to know what you want, such as when you want to start an outline within your document. You might have had the experience of having Microsoft Word suddenly determine, at the wrong moment, that you are creating an indented outline. While I am all for the automation of petty tasks, this is different.

From my point of view, this type of design feature is nonsense, since you end up having to work more than you would otherwise in order to manipulate the software's expectations of you. The real function of the feature isn't to make life easier for people. Instead, it promotes a new philosophy: that the computer is evolving into a life-form that can understand people better than people can understand themselves.

Another example is what I call the "race to be most meta." If a design like Facebook or Twitter depersonalizes people a little bit, then another service like Friendfeed—which may not even exist by the time this book is published—might soon come along to aggregate the previous layers of aggregation, making individual people even more abstract, and the illusion of high-level metaness more celebrated.

Information Doesn't Deserve to Be Free

"Information wants to be free." So goes the saying. Stewart Brand, the founder of the *Whole Earth Catalog*, seems to have said it first.

I say that information doesn't deserve to be free.

Cybernetic totalists love to think of the stuff as if it were alive and had its own ideas and ambitions. But what if information is inanimate? What if it's even less than inanimate, a mere artifact of human thought? What if only humans are real, and information is not?

Of course, there is a technical use of the term "information" that refers to something entirely real. This is the kind of information that's related to entropy. But that fundamental kind of information, which exists independently of the culture of an observer, is not the same as the kind we can put in computers, the kind that supposedly wants to be free.

Information is alienated experience.

You can think of culturally decodable information as a potential form of experience, very much as you can think of a brick resting on a ledge as storing potential energy. When the brick is prodded to fall, the energy is revealed. That is only possible because it was lifted into place at some point in the past.

In the same way, stored information might cause experience to be revealed if it is prodded in the right way. A file on a hard disk does indeed contain information of the kind that objectively exists. The fact that the

bits are discernible instead of being scrambled into mush—the way heat scrambles things—is what makes them bits.

But if the bits can potentially mean something to someone, they can only do so if they are experienced. When that happens, a commonality of culture is enacted between the storer and the retriever of the bits. Experience is the only process that can de-alienate information.

Information of the kind that purportedly wants to be free is nothing but a shadow of our own minds, and wants nothing on its own. It will not suffer if it doesn't get what it wants.

But if you want to make the transition from the old religion, where you hope God will give you an afterlife, to the new religion, where you hope to become immortal by getting uploaded into a computer, then you have to believe information is real and alive. So for you, it will be important to redesign human institutions like art, the economy, and the law to reinforce the perception that information is alive. You demand that the rest of us live in your new conception of a state religion. You need us to deify information to reinforce your faith.

The Apple Falls Again

It's a mistake with a remarkable origin. Alan Turing articulated it, just before his suicide.

Turing's suicide is a touchy subject in computer science circles. There's an aversion to talking about it much, because we don't want our founding father to seem like a tabloid celebrity, and we don't want his memory trivialized by the sensational aspects of his death.

The legacy of Turing the mathematician rises above any possible sensationalism. His contributions were supremely elegant and foundational. He gifted us with wild leaps of invention, including much of the mathematical underpinnings of digital computation. The highest award in computer science, our Nobel Prize, is named in his honor.

Turing the cultural figure must be acknowledged, however. The first thing to understand is that he was one of the great heroes of World War II. He was the first "cracker," a person who uses computers to defeat an enemy's security measures. He applied one of the first computers to break a Nazi secret code, called Enigma, which Nazi mathematicians

had believed was unbreakable. Enigma was decoded by the Nazis in the field using a mechanical device about the size of a cigar box. Turing reconceived it as a pattern of bits that could be analyzed in a computer, and cracked it wide open. Who knows what world we would be living in today if Turing had not succeeded?

The second thing to know about Turing is that he was gay at a time when it was illegal to be gay. British authorities, thinking they were doing the most compassionate thing, coerced him into a quack medical treatment that was supposed to correct his homosexuality. It consisted, bizarrely, of massive infusions of female hormones.

In order to understand how someone could have come up with that plan, you have to remember that before computers came along, the steam engine was a preferred metaphor for understanding human nature. All that sexual pressure was building up and causing the machine to malfunction, so the opposite essence, the female kind, ought to balance it out and reduce the pressure. This story should serve as a cautionary tale. The common use of computers, as we understand them today, as sources for models and metaphors of ourselves is probably about as reliable as the use of the steam engine was back then.

Turing developed breasts and other female characteristics and became terribly depressed. He committed suicide by lacing an apple with cyanide in his lab and eating it. Shortly before his death, he presented the world with a spiritual idea, which must be evaluated separately from his technical achievements. This is the famous Turing test. It is extremely rare for a genuinely new spiritual idea to appear, and it is yet another example of Turing's genius that he came up with one.

Turing presented his new offering in the form of a thought experiment, based on a popular Victorian parlor game. A man and a woman hide, and a judge is asked to determine which is which by relying only on the texts of notes passed back and forth.

Turing replaced the woman with a computer. Can the judge tell which is the man? If not, is the computer conscious? Intelligent? Does it deserve equal rights?

It's impossible for us to know what role the torture Turing was enduring at the time played in his formulation of the test. But it is undeniable that one of the key figures in the defeat of fascism was destroyed, by our

side, after the war, because he was gay. No wonder his imagination pondered the rights of strange creatures.

When Turing died, software was still in such an early state that no one knew what a mess it would inevitably become as it grew. Turing imagined a pristine, crystalline form of existence in the digital realm, and I can imagine it might have been a comfort to imagine a form of life apart from the torments of the body and the politics of sexuality. It's notable that it is the woman who is replaced by the computer, and that Turing's suicide echoes Eve's fall.

The Turing Test Cuts Both Ways

Whatever the motivation, Turing authored the first trope to support the idea that bits can be alive on their own, independent of human observers. This idea has since appeared in a thousand guises, from artificial intelligence to the hive mind, not to mention many overhyped Silicon Valley start-ups.

It seems to me, however, that the Turing test has been poorly interpreted by generations of technologists. It is usually presented to support the idea that machines can attain whatever quality it is that gives people consciousness. After all, if a machine fooled you into believing it was conscious, it would be bigoted for you to still claim it was not.

What the test really tells us, however, even if it's not necessarily what Turing hoped it would say, is that machine intelligence can only be known in a relative sense, in the eyes of a human beholder.*

The AI way of thinking is central to the ideas I'm criticizing in this

*One extension of the tragedy of Turing's death is that he didn't live long enough to articulate all that he probably would have about his own point of view on the Turing test.

Historian George Dyson suggests that Turing might have sided *against* the cybernetic totalists. For instance, here is an excerpt from a paper Turing wrote in 1939, titled "Systems of Logic Based on Ordinals": "We have been trying to see how far it is possible to eliminate intuition, and leave only ingenuity. We do not mind how much ingenuity is required, and therefore assume it to be available in unlimited supply." The implication seems to be that we are wrong to imagine that ingenuity can be infinite, even with computing clouds, so therefore intuition will never be made obsolete.

Turing's 1950 paper on the test includes this extraordinary passage: "In attempting to construct such machines we should not be irreverently usurping His power of creating souls, any more than we are in the procreation of children: rather we are, in either case, instruments of His will providing mansions for the souls that He creates."

book. If a machine can be conscious, then the computing cloud is going to be a better and far more capacious consciousness than is found in an individual person. If you believe this, then working for the benefit of the cloud over individual people puts you on the side of the angels.

But the Turing test cuts both ways. You can't tell if a machine has gotten smarter or if you've just lowered your own standards of intelligence to such a degree that the machine seems smart. If you can have a conversation with a simulated person presented by an AI program, can you tell how far you've let your sense of personhood degrade in order to make the illusion work for you?

People degrade themselves in order to make machines seem smart all the time. Before the crash, bankers believed in supposedly intelligent algorithms that could calculate credit risks before making bad loans. We ask teachers to teach to standardized tests so a student will look good to an algorithm. We have repeatedly demonstrated our species' bottomless ability to lower our standards to make information technology look good. Every instance of intelligence in a machine is ambiguous.

The same ambiguity that motivated dubious academic AI projects in the past has been repackaged as mass culture today. Did that search engine really know what you want, or are you playing along, lowering your standards to make it seem clever? While it's to be expected that the human perspective will be changed by encounters with profound new technologies, the exercise of treating machine intelligence as real requires people to reduce their mooring to reality.

A significant number of AI enthusiasts, after a protracted period of failed experiments in tasks like understanding natural language, eventually found consolation in the adoration for the hive mind, which yields better results because there are real people behind the curtain.

Wikipedia, for instance, works on what I call the Oracle illusion, in which knowledge of the human authorship of a text is suppressed in order to give the text superhuman validity. Traditional holy books work in precisely the same way and present many of the same problems.

This is another of the reasons I sometimes think of cybernetic totalist culture as a new religion. The designation is much more than an approximate metaphor, since it includes a new kind of quest for an afterlife. It's so weird to me that Ray Kurzweil wants the global computing cloud to

scoop up the contents of our brains so we can live forever in virtual reality. When my friends and I built the first virtual reality machines, the whole point was to make this world more creative, expressive, empathic, and interesting. It was not to escape it.

A parade of supposedly distinct "big ideas" that amount to the worship of the illusions of bits has enthralled Silicon Valley, Wall Street, and other centers of power. It might be Wikipedia or simulated people on the other end of the phone line. But really we are just hearing Turing's mistake repeated over and over.

Or Consider Chess

Will trendy cloud-based economics, science, or cultural processes outpace old-fashioned approaches that demand human understanding? No, because it is only encounters with human understanding that allow the contents of the cloud to exist.

Fragment liberation culture breathlessly awaits future triumphs of technology that will bring about the Singularity or other imaginary events. But there are already a few examples of how the Turing test has been approximately passed, and has reduced personhood. Chess is one.

The game of chess possesses a rare combination of qualities: it is easy to understand the rules, but it is hard to play well; and, most important, the urge to master it seems timeless. Human players achieve ever higher levels of skill, yet no one will claim that the quest is over.

Computers and chess share a common ancestry. Both originated as tools of war. Chess began as a battle simulation, a mental martial art. The design of chess reverberates even further into the past than that—all the way back to our sad animal ancestry of pecking orders and competing clans.

Likewise, modern computers were developed to guide missiles and break secret military codes. Chess and computers are both direct descendants of the violence that drives evolution in the natural world, however sanitized and abstracted they may be in the context of civilization. The drive to compete is palpable in both computer science and chess, and when they are brought together, adrenaline flows.

What makes chess fascinating to computer scientists is precisely that

we're bad at it. From our point of view, human brains routinely do things that seem almost insuperably difficult, like understanding sentences—yet we don't hold sentence-comprehension tournaments, because we find that task too easy, too ordinary.

Computers fascinate and frustrate us in a similar way. Children can learn to program them, yet it is extremely difficult for even the most accomplished professional to program them well. Despite the evident potential of computers, we know full well that we have not thought of the best programs to write.

But all of this is not enough to explain the outpouring of public angst on the occasion of Deep Blue's victory in May 1997 over world chess champion Gary Kasparov, just as the web was having its first major influences on popular culture. Regardless of all the old-media hype, it was clear that the public's response was genuine and deeply felt. For millennia, mastery of chess had indicated the highest, most refined intelligence—and now a computer could play better than the very best human.

There was much talk about whether human beings were still special, whether computers were becoming our equal. By now, this sort of thing wouldn't be news, since people have had the AI way of thinking pounded into their heads so much that it is sounding like believable old news. The AI way of framing the event was unfortunate, however. What happened was primarily that a team of computer scientists built a very fast machine and figured out a better way to represent the problem of how to choose the next move in a chess game. People, not machines, performed this accomplishment.

The Deep Blue team's central victory was one of clarity and elegance of thought. In order for a computer to beat the human chess champion, two kinds of progress had to converge: an increase in raw hardware power and an improvement in the sophistication and clarity with which the decisions of chess play are represented in software. This dual path made it hard to predict the year, but not the eventuality, that a computer would triumph.

If the Deep Blue team had not been as good at the software problem, a computer would still have become the world champion at some later date, thanks to sheer brawn. So the suspense lay in wondering not whether a chess-playing computer would ever beat the best human chess

player, but to what degree programming elegance would play a role in the victory. Deep Blue won earlier than it might have, scoring a point for elegance.

The public reaction to the defeat of Kasparov left the computer science community with an important question, however. Is it useful to portray computers themselves as intelligent or humanlike in any way? Does this presentation serve to clarify or to obscure the role of computers in our lives?

Whenever a computer is imagined to be intelligent, what is really happening is that humans have abandoned aspects of the subject at hand in order to remove from consideration whatever the computer is blind to. This happened to chess itself in the case of the Deep Blue–Kasparov tournament.

There is an aspect of chess that is a little like poker—the staring down of an opponent, the projection of confidence. Even though it is relatively easier to write a program to "play" poker than to play chess, poker is really a game centering on the subtleties of nonverbal communication between people, such as bluffing, hiding emotion, understanding your opponents' psychologies, and knowing how to bet accordingly. In the wake of Deep Blue's victory, the poker side of chess has been largely overshadowed by the abstract, algorithmic aspect—while, ironically, it was in the poker side of the game that Kasparov failed critically.

Kasparov seems to have allowed himself to be spooked by the computer, even after he had demonstrated an ability to defeat it on occasion. He might very well have won if he had been playing a human player with exactly the same move-choosing skills as Deep Blue (or at least as Deep Blue existed in 1997). Instead, Kasparov detected a sinister stone face where in fact there was absolutely nothing. While the contest was not intended as a Turing test, it ended up as one, and Kasparov was fooled.

As I pointed out earlier, the idea of AI has shifted the psychological projection of adorable qualities from computer programs alone to a different target: computer-plus-crowd constructions. So, in 1999 a wikilike crowd of people, including chess champions, gathered to play Kasparov in an online game called "Kasparov versus the World." In this case Kasparov won, though many believe that it was only because of backstabbing between members of the crowd. We technologists are cease-

lessly intrigued by rituals in which we attempt to pretend that people are obsolete.

The attribution of intelligence to machines, crowds of fragments, or other nerd deities obscures more than it illuminates. When people are told that a computer is intelligent, they become prone to changing themselves in order to make the computer appear to work better, instead of demanding that the computer be changed to become more useful. People already tend to defer to computers, blaming themselves when a digital gadget or online service is hard to use.

Treating computers as intelligent, autonomous entities ends up standing the process of engineering on its head. We can't afford to respect our own designs so much.

The Circle of Empathy

The most important thing to ask about any technology is how it changes people. And in order to ask that question I've used a mental device called the "circle of empathy" for many years. Maybe you'll find it useful as well. (The Princeton philosopher often associated with animal rights, Peter Singer, uses a similar term and idea, seemingly a coincident coinage.)

An imaginary circle of empathy is drawn by each person. It circumscribes the person at some distance, and corresponds to those things in the world that deserve empathy. I like the term "empathy" because it has spiritual overtones. A term like "sympathy" or "allegiance" *might* be more precise, but I want the chosen term to be slightly mystical, to suggest that we might not be able to fully understand what goes on between us and others, that we should leave open the possibility that the relationship can't be represented in a digital database.

If someone falls within your circle of empathy, you wouldn't want to see him or her killed. Something that is clearly outside the circle is fair game. For instance, most people would place all other people within the circle, but most of us are willing to see bacteria killed when we brush our teeth, and certainly don't worry when we see an inanimate rock tossed aside to keep a trail clear.

The tricky part is that some entities reside close to the edge of the circle. The deepest controversies often involve whether something or someone should lie just inside or just outside the circle. For instance, the idea of slavery depends on the placement of the slave outside the circle, to make some people nonhuman. Widening the circle to include all people and end slavery has been one of the epic strands of the human story—and it isn't quite over yet.

A great many other controversies fit well in the model. The fight over abortion asks whether a fetus or embryo should be in the circle or not, and the animal rights debate asks the same about animals.

When you change the contents of your circle, you change your conception of yourself. The center of the circle shifts as its perimeter is changed. The liberal impulse is to expand the circle, while conservatives tend to want to restrain or even contract the circle.

Empathy Inflation and Metaphysical Ambiguity

Are there any legitimate reasons not to expand the circle as much as possible? There are.

To expand the circle indefinitely can lead to oppression, because the rights of potential entities (as perceived by only some people) can conflict with the rights of indisputably real people. An obvious example of this is found in the abortion debate. If outlawing abortions did not involve commandeering control of the bodies of other people (pregnant women, in this case), then there wouldn't be much controversy. We would find an easy accommodation.

Empathy inflation can also lead to the lesser, but still substantial, evils of incompetence, trivialization, dishonesty, and narcissism. You cannot live, for example, without killing bacteria. Wouldn't you be projecting your own fantasies on single-cell organisms that would be indifferent to them at best? Doesn't it really become about you instead of the cause at that point? Do you go around blowing up other people's toothbrushes? Do you think the bacteria you saved are morally equivalent to former slaves—and if you do, haven't you diminished the status of those human beings? Even if you can follow your passion to free and protect the world's bacteria with a pure heart, haven't you divorced yourself from

the reality of interdependence and transience of all things? You can try to avoid killing bacteria on special occasions, but you need to kill them to live. And even if you are willing to die for your cause, you can't prevent bacteria from devouring your own body when you die.

Obviously the example of bacteria is extreme, but it shows that the circle is only meaningful if it is finite. If we lose the finitude, we lose our own center and identity. The fable of the Bacteria Liberation Front can serve as a parody of any number of extremist movements on the left or the right.

At the same time, I have to admit that I find it impossible to come to a definitive position on many of the most familiar controversies. I am all for animal rights, for instance, but only as a hypocrite. I eat chicken, but I can't eat cephalopods—octopus and squid—because I admire their neurological evolution so intensely. (Cephalopods also suggest an alternate way to think about the long-term future of technology that avoids certain moral dilemmas—something I'll explain later in the book.)

How do I draw my circle? I just spend time with the various species and decide if they feel like they are in my circle or not. I've raised chickens and somehow haven't felt empathy toward them. They are little more than feathery servo-controlled mechanisms compared to goats, for instance, which I have also raised, and will not eat. On the other hand, a colleague of mine, virtual reality researcher Adrian Cheok, feels such empathy with chickens that he built teleimmersion suits for them so that he could telecuddle them from work. We all have to live with our imperfect ability to discern the proper boundaries of our circles of empathy. There will always be cases where reasonable people will disagree. I don't go around telling other people not to eat cephalopods or goats.

The border between person and nonperson might be found somewhere in the embryonic sequence from conception to baby, or in the development of the young child, or the teenager. Or it might be best defined in the phylogenetic path from ape to early human, or perhaps in the cultural history of ancient peasants leading to modern citizens. It might exist somewhere in a continuum between small and large computers. It might have to do with which thoughts you have; maybe self-reflective thoughts or the moral capacity for empathy makes you human. These are some of the many gates to personhood that have been proposed, but none of them seem definitive to me. The borders of personhood remain variegated and fuzzy.

Paring the Circle

Just because we are unable to know precisely where the circle of empathy should lie does not mean that we are unable to know anything at all about it. If we are only able to be approximately moral, that doesn't mean we should give up trying to be moral at all. The term "morality" is usually used to describe our treatment of others, but in this case I am applying it to ourselves just as much.

The dominant open digital culture places digital information processing in the role of the embryo as understood by the religious right, or the bacteria in my reductio ad absurdum fable. The error is classical, but the consequences are new. I fear that we are beginning to design ourselves to suit digital models of us, and I worry about a leaching of empathy and humanity in that process.

The rights of embryos are based on extrapolation, while the rights of a competent adult person are as demonstrable as anything can be, since people speak for themselves. There are plenty of examples where it's hard to decide where to place faith in personhood because a proposed being, while it might be deserving of empathy, cannot speak for itself.

Should animals have the same rights as humans? There are special perils when some people hear voices, and extend empathy, that others do not. If it's at all possible, these are exactly the situations that must be left to people close to a given situation, because otherwise we'll ruin personal freedom by enforcing metaphysical ideas on one another.

In the case of slavery, it turned out that, given a chance, slaves could not just speak for themselves, they could speak intensely and well. Moses was unambiguously a person. Descendants of more recent slaves, like Martin Luther King Jr., demonstrated transcendent eloquence and empathy.

The new twist in Silicon Valley is that some people—very influential people—believe they are hearing algorithms and crowds and other internet-supported nonhuman entities speak for themselves. I don't hear those voices, though—and I believe those who do are fooling themselves.

Thought Experiments: The Ship of Theseus Meets the Infinite Library of Borges

To help you learn to doubt the fantasies of the cybernetic totalists, I offer two dueling thought experiments.

The first one has been around a long time. As Daniel Dennett tells it: Imagine a computer program that can simulate a neuron, or even a network of neurons. (Such programs have existed for years and in fact are getting quite good.) Now imagine a tiny wireless device that can send and receive signals to neurons in the brain. Crude devices a little like this already exist; years ago I helped Joe Rosen, a reconstructive plastic surgeon at Dartmouth Medical School, build one—the "nerve chip," which was an early attempt to route around nerve damage using prosthetics.

To get the thought experiment going, hire a neurosurgeon to open your skull. If that's an inconvenience, swallow a nano-robot that can perform neurosurgery. Replace one nerve in your brain with one of those wireless gadgets. (Even if such gadgets were already perfected, connecting them would not be possible today. The artificial neuron would have to engage all the same synapses—around seven thousand, on average—as the biological nerve it replaced.)

Next, the artificial neuron will be connected over a wireless link to a simulation of a neuron in a nearby computer. Every neuron has unique chemical and structural characteristics that must be included in the program. Do the same with your remaining neurons. There are between 100 billion and 200 billion neurons in a human brain, so even at only a second per neuron, this will require tens of thousands of years.

Now for the big question: Are you still conscious after the process has been completed?

Furthermore, because the computer is completely responsible for the dynamics of your brain, you can forgo the physical artificial neurons and let the neuron-control programs connect with one another through software alone. Does the computer then become a person? If you believe in consciousness, is your consciousness now in the computer, or perhaps in the software? The same question can be asked about souls, if you believe in them.

Bigger Borges

Here's a second thought experiment. It addresses the same question from the opposite angle. Instead of changing the program running on the computer, it changes the design of the computer.

First, imagine a marvelous technology: an array of flying laser scanners that can measure the trajectories of all the hailstones in a storm. The scanners send all the trajectory information to your computer via a wireless link.

What would anyone do with this data? As luck would have it, there's a wonderfully geeky store in this thought experiment called the Ultimate Computer Store, which sells a great many designs of computers. In fact, every possible computer design that has fewer than some really large number of logic gates is kept in stock.

You arrive at the Ultimate Computer Store with a program in hand. A salesperson gives you a shopping cart, and you start trying out your program on various computers as you wander the aisles. Once in a while you're lucky, and the program you brought from home will run for a reasonable period of time without crashing on a computer. When that happens, you drop the computer in the shopping cart.

For a program, you could even use the hailstorm data. Recall that a computer program is nothing but a list of numbers; there must be some computers in the Ultimate Computer Store that will run it! The strange thing is that each time you find a computer that runs the hailstorm data as a program, the program does something different.

After a while, you end up with a few million word processors, some amazing video games, and some tax-preparation software—all the same program, as it runs on different computer designs. This takes time; in the real world the universe probably wouldn't support conditions for life long enough for you to make a purchase. But this is a thought experiment, so don't be picky.

The rest is easy. Once your shopping cart is filled with a lot of computers that run the hailstorm data, settle down in the store's café. Set up the computer from the first thought experiment, the one that's running a copy of your brain. Now go through all your computers and compare what each one does with what the computer from the first experiment

does. Do this until you find a computer that runs the hailstorm data as a program equivalent to your brain.

How do you know when you've found a match? There are endless options. For mathematical reasons, you can never be absolutely sure of what a big program does or if it will crash, but if you found a way to be satisfied with the software neuron replacements in the first thought experiment, you have already chosen your method to approximately evaluate a big program. Or you could even find a computer in your cart that interprets the motion of the hailstorm over an arbitrary period of time as equivalent to the activity of the brain program over a period of time. That way, the dynamics of the hailstorm are matched to the brain program beyond just one moment in time.

After you've done all this, is the hailstorm now conscious? Does it have a soul?

The Metaphysical Shell Game

The alternative to sprinkling magic dust on people is sprinkling it on computers, the hive mind, the cloud, the algorithm, or some other cybernetic object. The right question to ask is, Which choice is crazier?

If you try to pretend to be certain that there's no mystery in something like consciousness, the mystery that is there can pop out elsewhere in an inconvenient way and ruin your objectivity as a scientist. You enter into a metaphysical shell game that can make you dizzy. For instance, you can propose that consciousness is an illusion, but by definition consciousness is the one thing that isn't reduced if it is an illusion.

There's a way that consciousness and time are bound together. If you try to remove any potential hint of mysteriousness from consciousness, you end up mystifying time in an absurd way.

Consciousness is situated in time, because you can't experience a lack of time, and you can't experience the future. If consciousness isn't anything but a false thought in the computer that is your brain, or the universe, then what exactly *is* it that is situated in time? The present moment, the only other thing that could be situated in time, must in that case be a freestanding object, independent of the way it is experienced.

The present moment is a rough concept, from a scientific point of view, because of relativity and the latency of thoughts moving in the brain. We have no means of defining either a single global physical present moment or a precise cognitive present moment. Nonetheless, there must be *some* anchor, perhaps a very fuzzy one, somewhere, somehow, for it to be possible to even speak of it.

Maybe you could imagine the present moment as a metaphysical marker traveling through a timeless version of reality, in which the past and the future are already frozen in place, like a recording head moving across a hard disk.

If you are certain the experience of time is an illusion, all you have left is time itself. *Something* has to be situated—in a kind of metatime or something—in order for the illusion of the present moment to take place at all. You force yourself to say that time itself travels through reality. This is an absurd, circular thought.

To call consciousness an illusion is to give time a supernatural quality—maybe some kind of spooky nondeterminism. Or you can choose a different shell in the game and say that time is natural (not supernatural), and that the present moment is only a possible concept because of consciousness.

The mysterious stuff can be shuffled around, but it is best to just admit when some trace of mystery remains, in order to be able to speak as clearly as possible about the many things that can actually be studied or engineered methodically.

I acknowledge that there are dangers when you allow for the legitimacy of a metaphysical idea (like the potential for consciousness to be something beyond computation). No matter how careful you are not to "fill in" the mystery with superstitions, you might encourage some fundamentalists or new-age romantics to cling to weird beliefs. "Some dreadlocked computer scientist says consciousness might be more than a computer? Then my food supplement must work!"

But the danger of an engineer pretending to know more than he really does is the greater danger, especially when he can reinforce the illusion through the use of computation. The cybernetic totalists awaiting the Singularity are nuttier than the folks with the food supplements.

The Zombie Army

Do fundamental metaphysical—or supposedly antimetaphysical—beliefs trickle down into the practical aspects of our thinking or our personalities? They do. They can turn a person into what philosophers call a "zombie."

Zombies are familiar characters in philosophical thought experiments. They are like people in every way except that they have no internal experience. They are unconscious, but give no externally measurable evidence of that fact. Zombies have played a distinguished role as fodder in the rhetoric used to discuss the mind-body problem and consciousness research. There has been much debate about whether a true zombie could exist, or if internal subjective experience inevitably colors either outward behavior or measurable events in the brain in some way.

I claim that there is one measurable difference between a zombie and a person: a zombie has a different philosophy. Therefore, zombies can only be detected if they happen to be professional philosophers. A philosopher like Daniel Dennett is obviously a zombie.

Zombies and the rest of us do not have a symmetrical relationship. Unfortunately, it is only possible for nonzombies to observe the telltale sign of zombiehood. To zombies, everyone looks the same.

If there are enough zombies recruited into our world, I worry about the potential for a self-fulfilling prophecy. Maybe if people pretend they are not conscious or do not have free will—or that the cloud of online people is a person; if they pretend there is nothing special about the perspective of the individual—then perhaps we have the power to make it so. We might be able to collectively achieve antimagic.

Humans are free. We can commit suicide for the benefit of a Singularity. We can engineer our genes to better support an imaginary hive mind. We can make culture and journalism into second-rate activities and spend centuries remixing the detritus of the 1960s and other eras from before individual creativity went out of fashion.

Or we can believe in ourselves. By chance, it might turn out we are real.