PART 3

Technology's Impact on Society

Although barely one year old, the 21st century is already promising scientific and technological innovations that were previously the venue of science fiction writers. Developments in nanotechnology, MicroElectroMechanical Systems (MEMS), genetic engineering, cloning, robotics, and optical and quantum computing are likely to have a great impact upon the future structure of our societies. While there is wide dispute as to whether these advances will be a boon or a bane to humanity, there is a consensus opinion that change is coming. In Part 3, four papers explore the debate and offer suggestions as to how we should prepare for the coming storm.

Chapter 3 contains a reprint of Bill Joy's controversial and thought provoking *Wired* article, "Why the Future Doesn't Need Us." Joy, CEO, chief scientist and co-founder of Sun Microsystems, sees a future that is frightfully bleak. "Accustomed to living with almost routine scientific breakthroughs, we have yet to come to terms with the fact that the most compelling 21st-century technologies—robotics, genetic engineering, and nanotechnology—pose a different threat than the technologies that have come before." He warns that we have become so enamored of the quest for scientific knowledge, that we no longer bother to consider the impacts of our discoveries. Moreover, Joy is concerned that the self-replicating abilities of novel innovations in genetic engineering, robotics, and nanotechnology (GNR) could produce dire results, including the end of humanity.

In light of this potential danger, Joy suggests a controversial solution: verifiable, voluntary relinquishment of dangerous GNR technologies. He writes, "In the 21st century, this requires vigilance and personal responsibility by those who would work on both NBC [nuclear, biological, and chemical weapons] and GNR technologies to avoid implementing weapons of mass destruction and knowledge-enabled mass destruction."

Joy's view is rebutted in Chapter 4, in a reprint of an *Industry Standard* article by John Seely Brown of the Xerox Palo Alto Research Center, and Paul Duguid of the University of California at Berkeley. In their paper, the authors explain why they believe Joy's vision of the future is just plain wrong. They use the ballyhoo associated with the dawning of the nuclear age in the 1950s to make their case, by reminding readers that technoenthusiasts had previously predicted the "end of power monopolies, the emergence of the 'electronic cottage,' the death of the city and the decline of the corporation." The authors believe much of Joy's pessimism stems from a tunnel vision that "leaves people out of the picture and focuses on technology in splendid isolation." Brown and Duguid hold that the future is only profoundly bleak if one looks at it through the "6-D" lenses of: demassification, decentralization, disintermediation, despacialization, disaggregation and demarketization.

Michael L. Dertouzos' article, "Not by Reason Alone," originally published in *Technology Review* is reprinted in Chapter 5. Dertouzos is also critical of Joy's vision. He is troubled by "the arrogant notion that human logic can anticipate the effects of intended or unintended acts, and the more arrogant notion that human reasoning can determine the course of the universe." He counters that we are limited in our ability to assess consequences, and suggests that we rely on our humanity, feelings and beliefs when determining the impacts of technology. Unlike Joy, who would prohibit certain types of research because of their potential dangers, Dertouzos states that we should instead proceed, but "stay vigilant, ready to stop, when danger is imminent, using our full humanity to make that determination."

Part 3 concludes with a paper by Michael M. Crow of Columbia University, and Daniel Sarewitz of the Center for Science, Policy, and Outcomes. They see technology and innovation as forces that can remake social structures and create new phenomena, which in turn, lead to new institutions and response mechanisms. Crow and Sarewitz believe that nanotechnology has the potential to be another catalyst for sweeping societal change. Their point however is "not to predict the future of nanotechnology and its impacts—an impossible task—but to illustrate the direction and scale of thinking that will be necessary if we are to successfully manage the interaction of new knowledge and innovation with society." They warn that the revolution is coming, and we can allow ourselves to be caught by surprise, or we can prepare for it "in order to enhance the benefits, and reduce the disruption and dislocation, that must accompany any revolution."

3 Why the Future Doesn't Need Us

Bill Joy

From the moment I became involved in the creation of new technologies, their ethical dimensions have concerned me, but it was only in the autumn of 1998 that I became anxiously aware of how great are the dangers facing us in the 21st century. I can date the onset of my unease to the day I met Ray Kurzweil, the deservedly famous inventor of the first reading machine for the blind and many other amazing things.

Ray and I were both speakers at George Gilder's Telecosm conference, and I encountered him by chance in the bar of the hotel after both our sessions were over. I was sitting with John Searle, a Berkeley philosopher who studies consciousness. While we were talking, Ray approached and a conversation began, the subject of which haunts me to this day.

I had missed Ray's talk and the subsequent panel that Ray and John had been on, and they now picked right up where they'd left off, with Ray saying that the rate of improvement of technology was going to accelerate and that we were going to become robots or fuse with robots or something like that, and John countering that this couldn't happen, because the robots couldn't be conscious.

While I had heard such talk before, I had always felt sentient robots were in the realm of science fiction. But now, from someone I respected, I was hearing a strong argument that they were a near-term possibility. I was taken aback, especially given Ray's proven ability to imagine and create the future. I already knew that new technologies like genetic engineering and nanotechnology were giving us the power to remake the world, but a realistic and imminent scenario for intelligent robots surprised me.

Bill Joy is co-founder, chief scientist, and corporate executive officer of Sun Microsystems. This article is reprinted with permission from Wired *8.04, April 2000.*

It's easy to get jaded about such breakthroughs. We hear in the news almost every day of some kind of technological or scientific advance. Yet this was no ordinary prediction. In the hotel bar, Ray gave me a partial preprint of his then-forthcoming book *The Age of Spiritual Machines,* which outlined a utopia he foresaw—one in which humans gained near immortality by becoming one with robotic technology. On reading it, my sense of unease only intensified; I felt sure he had to be understating the dangers, understating the probability of a bad outcome along this path.

I found myself most troubled by a passage detailing a dystopian scenario:

The New Luddite Challenge

> First let us postulate that the computer scientists succeed in developing intelligent machines that can do all things better than human beings can do them. In that case presumably all work will be done by vast, highly organized systems of machines and no human effort will be necessary. Either of two cases might occur. The machines might be permitted to make all of their own decisions without human oversight, or else human control over the machines might be retained.
>
> If the machines are permitted to make all their own decisions, we can't make any conjectures as to the results, because it is impossible to guess how such machines might behave. We only point out that the fate of the human race would be at the mercy of the machines. It might be argued that the human race would never be foolish enough to hand over all the power to the machines. But we are suggesting neither that the human race would voluntarily turn power over to the machines nor that the machines would willfully seize power. What we do suggest is that the human race might easily permit itself to drift into a position of such dependence on the machines that it would have no practical choice but to accept all of the machines' decisions. As society and the problems that face it become more and more complex and machines become more and more intelligent, people will let machines make more of their decisions for them, simply because machine-made decisions will bring better results than man-made ones. Eventually a stage may

> be reached at which the decisions necessary to keep the system running will be so complex that human beings will be incapable of making them intelligently. At that stage the machines will be in effective control. People won't be able to just turn the machines off, because they will be so dependent on them that turning them off would amount to suicide.
>
> On the other hand it is possible that human control over the machines may be retained. In that case the average man may have control over certain private machines of his own, such as his car or his personal computer, but control over large systems of machines will be in the hands of a tiny elite—just as it is today, but with two differences. Due to improved techniques the elite will have greater control over the masses; and because human work will no longer be necessary the masses will be superfluous, a useless burden on the system. If the elite is ruthless they may simply decide to exterminate the mass of humanity. If they are humane they may use propaganda or other psychological or biological techniques to reduce the birth rate until the mass of humanity becomes extinct, leaving the world to the elite. Or, if the elite consists of soft-hearted liberals, they may decide to play the role of good shepherds to the rest of the human race. They will see to it that everyone's physical needs are satisfied, that all children are raised under psychologically hygienic conditions, that everyone has a wholesome hobby to keep him busy, and that anyone who may become dissatisfied undergoes "treatment" to cure his "problem." Of course, life will be so purposeless that people will have to be biologically or psychologically engineered either to remove their need for the power process or make them "sublimate" their drive for power into some harmless hobby. These engineered human beings may be happy in such a society, but they will most certainly not be free. They will have been reduced to the status of domestic animals.[1]

In the book, you don't discover until you turn the page that the author of this passage is Theodore Kaczynski—the Unabomber. I am no apologist for Kaczynski. His bombs killed three people during a 17-year terror campaign and wounded many others. One of his bombs gravely injured my friend David Gelernter, one of the most brilliant and visionary computer scientists of our time. Like many of my colleagues, I felt that I could easily have been the Unabomber's next target.

Kaczynski's actions were murderous and, in my view, criminally insane. He is clearly a Luddite, but simply saying this does not dismiss his argument; as difficult as it is for me to acknowledge, I saw some merit in the reasoning in this single passage. I felt compelled to confront it.

Kaczynski's dystopian vision describes unintended consequences, a well-known problem with the design and use of technology, and one that is clearly related to Murphy's law—"Anything that can go wrong, will." (Actually, this is Finagle's law, which in itself shows that Finagle was right.) Our overuse of antibiotics has led to what may be the biggest such problem so far: the emergence of antibiotic-resistant and much more dangerous bacteria. Similar things happened when attempts to eliminate malarial mosquitoes using DDT caused them to acquire DDT resistance; malarial parasites likewise acquired multi-drug-resistant genes.[2]

The cause of many such surprises seems clear: The systems involved are complex, involving interaction among and feedback between many parts. Any changes to such a system will cascade in ways that are difficult to predict; this is especially true when human actions are involved.

I started showing friends the Kaczynski quote from *The Age of Spiritual Machines;* I would hand them Kurzweil's book, let them read the quote, and then watch their reaction as they discovered who had written it. At around the same time, I found Hans Moravec's book *Robot: Mere Machine to Transcendent Mind.* Moravec is one of the leaders in robotics research, and was a founder of the world's largest robotics research program, at Carnegie Mellon University. *Robot* gave me more material to try out on my friends—material surprisingly supportive of Kaczynski's argument. For example:

The Short Run (Early 2000s)

> Biological species almost never survive encounters with superior competitors. Ten million years ago, South and North America were separated by a sunken Panama isthmus. South America, like Australia today, was populated by marsupial mammals, including pouched equivalents of rats, deers, and tigers. When the isthmus connecting North and South America rose, it took only a few thousand years for the northern placental species, with slightly more effective metabolisms and reproductive and nervous systems, to displace and eliminate almost all the southern marsupials.

> In a completely free marketplace, superior robots would surely affect humans as North American placentals affected South American marsupials (and as humans have affected countless species). Robotic industries would compete vigorously among themselves for matter, energy, and space, incidentally driving their price beyond human reach. Unable to afford the necessities of life, biological humans would be squeezed out of existence.
>
> There is probably some breathing room, because we do not live in a completely free marketplace. Government coerces nonmarket behavior, especially by collecting taxes. Judiciously applied, governmental coercion could support human populations in high style on the fruits of robot labor, perhaps for a long while.

A textbook dystopia—and Moravec is just getting wound up. He goes on to discuss how our main job in the 21st century will be "ensuring continued cooperation from the robot industries" by passing laws decreeing that they be "nice,"[3] and to describe how seriously dangerous a human can be "once transformed into an unbounded superintelligent robot." Moravec's view is that the robots will eventually succeed us—that humans clearly face extinction.

I decided it was time to talk to my friend Danny Hillis. Danny became famous as the cofounder of Thinking Machines Corporation, which built a very powerful parallel supercomputer. Despite my current job title of Chief Scientist at Sun Microsystems, I am more a computer architect than a scientist, and I respect Danny's knowledge of the information and physical sciences more than that of any other single person I know. Danny is also a highly regarded futurist who thinks long-term—four years ago he started the Long Now Foundation, which is building a clock designed to last 10,000 years, in an attempt to draw attention to the pitifully short attention span of our society. (See "Test of Time," *Wired* 8.03, page 78.)

So I flew to Los Angeles for the express purpose of having dinner with Danny and his wife, Pati. I went through my now-familiar routine, trotting out the ideas and passages that I found so disturbing. Danny's answer—directed specifically at Kurzweil's scenario of humans merging with robots—came swiftly, and quite surprised me. He said, simply, that the changes would come gradually, and that we would get used to them.

But I guess I wasn't totally surprised. I had seen a quote from Danny in Kurzweil's book in which he said, "I'm as fond of my body as anyone, but if I can be 200 with a body of silicon, I'll take it." It seemed that he was at peace with this process and its attendant risks, while I was not.

While talking and thinking about Kurzweil, Kaczynski, and Moravec, I suddenly remembered a novel I had read almost 20 years ago—*The White Plague,* by Frank Herbert—in which a molecular biologist is driven insane by the senseless murder of his family. To seek revenge he constructs and disseminates a new and highly contagious plague that kills widely but selectively. (We're lucky Kaczynski was a mathematician, not a molecular biologist.) I was also reminded of the Borg of *Star Trek,* a hive of partly biological, partly robotic creatures with a strong destructive streak. Borg-like disasters are a staple of science fiction, so why hadn't I been more concerned about such robotic dystopias earlier? Why weren't other people more concerned about these nightmarish scenarios?

Part of the answer certainly lies in our attitude toward the new—in our bias toward instant familiarity and unquestioning acceptance. Accustomed to living with almost routine scientific breakthroughs, we have yet to come to terms with the fact that the most compelling 21st-century technologies—robotics, genetic engineering, and nanotechnology—pose a different threat than the technologies that have come before. Specifically, robots, engineered organisms, and nanobots share a dangerous amplifying factor: They can self-replicate. A bomb is blown up only once—but one bot can become many, and quickly get out of control.

Much of my work over the past 25 years has been on computer networking, where the sending and receiving of messages creates the opportunity for out-of-control replication. But while replication in a computer or a computer network can be a nuisance, at worst it disables a machine or takes down a network or network service. Uncontrolled self-replication in these newer technologies runs a much greater risk: a risk of substantial damage in the physical world.

Each of these technologies also offers untold promise: The vision of near immortality that Kurzweil sees in his robot dreams drives us forward; genetic engineering may soon provide treatments, if not outright cures, for most diseases; and nanotechnology and nanomedicine can address yet more ills. Together they could significantly extend our average life span and improve the quality of our lives. Yet, with each of these technologies, a sequence of small, individually sensible advances leads to an accumulation of great power and, concomitantly, great danger.

What was different in the 20th century? Certainly, the technologies underlying the weapons of mass destruction (WMD)—nuclear, biological, and chemical (NBC)—were powerful, and the weapons an enormous threat. But building nuclear weapons required, at least for a time, access to both rare—indeed, effectively unavailable—raw materials and highly protected information; biological and chemical weapons programs also tended to require large-scale activities.

The 21st-century technologies—genetics, nanotechnology, and robotics (GNR)—are so powerful that they can spawn whole new classes of accidents and abuses. Most dangerously, for the first time, these accidents and abuses are widely within the reach of individuals or small groups. They will not require large facilities or rare raw materials. Knowledge alone will enable the use of them.

Thus we have the possibility not just of weapons of mass destruction but of knowledge-enabled mass destruction (KMD), this destructiveness hugely amplified by the power of self-replication.

I think it is no exaggeration to say we are on the cusp of the further perfection of extreme evil, an evil whose possibility spreads well beyond that which weapons of mass destruction bequeathed to the nation-states, on to a surprising and terrible empowerment of extreme individuals.

Nothing about the way I got involved with computers suggested to me that I was going to be facing these kinds of issues.

My life has been driven by a deep need to ask questions and find answers. When I was 3, I was already reading, so my father took me to the elementary school, where I sat on the principal's lap and read him a story. I started school early, later skipped a grade, and escaped into books—I was incredibly motivated to learn. I asked lots of questions, often driving adults to distraction.

As a teenager I was very interested in science and technology. I wanted to be a ham radio operator but didn't have the money to buy the equipment. Ham radio was the Internet of its time: very addictive, and quite solitary. Money issues aside, my mother put her foot down—I was not to be a ham; I was antisocial enough already.

I may not have had many close friends, but I was awash in ideas. By high school, I had discovered the great science fiction writers. I remember especially Heinlein's *Have Spacesuit Will Travel* and Asimov's *I, Robot,* with its Three Laws of Robotics. I was enchanted by the descriptions of space travel, and wanted to have a telescope to look at the stars; since I had no money to buy or make one, I checked books on telescope-making out of the library and read about making them instead. I soared in my imagination.

Thursday nights my parents went bowling, and we kids stayed home alone. It was the night of Gene Roddenberry's original *Star Trek,* and the program made a big impression on me. I came to accept its notion that humans had a future in space, Western-style, with big heroes and adventures. Roddenberry's vision of the centuries to come was one with strong moral values, embodied in codes like the Prime Directive: to not interfere in the development of less technologically advanced civilizations. This had an incredible appeal to me; ethical humans, not robots, dominated this future, and I took Roddenberry's dream as part of my own.

I excelled in mathematics in high school, and when I went to the University of Michigan as an undergraduate engineering student I took the advanced curriculum of the mathematics majors. Solving math problems was an exciting challenge, but when I discovered computers I found something much more interesting: a machine into which you could put a program that attempted to solve a problem, after which the machine quickly checked the solution. The computer had a clear notion of correct and incorrect, true and false. Were my ideas correct? The machine could tell me. This was very seductive.

I was lucky enough to get a job programming early supercomputers and discovered the amazing power of large machines to numerically simulate advanced designs. When I went to graduate school at UC Berkeley in the mid-1970s, I started staying up late, often all night, inventing new worlds inside the machines. Solving problems. Writing the code that argued so strongly to be written.

In *The Agony and the Ecstasy,* Irving Stone's biographical novel of Michelangelo, Stone described vividly how Michelangelo released the statues from the stone, "breaking the marble spell," carving from the images in his mind.[4] In my most ecstatic moments, the software in the computer emerged in the same way. Once I had imagined it in my mind I felt that it was already there in the machine, waiting to be released. Staying up all night seemed a small price to pay to free it—to give the ideas concrete form.

After a few years at Berkeley I started to send out some of the software I had written—an instructional Pascal system, Unix utilities, and a text editor called vi (which is still, to my surprise, widely used more than 20 years later)—to others who had similar small PDP-11 and VAX minicomputers. These adventures in software eventually turned into the Berkeley version of the Unix operating system, which became a personal "success disaster"—so many people wanted it that I never finished my Ph.D. Instead I got a job working for DARPA putting Berkeley Unix on the Internet and fixing it to be reliable and to run large research applications well. This was all great fun and very rewarding. And, frankly, I saw no robots here, or anywhere near.

Still, by the early 1980s, I was drowning. The Unix releases were very successful, and my little project of one soon had money and some staff, but the problem at Berkeley was always office space rather than money—there wasn't room for the help the project needed, so when the other founders of Sun Microsystems showed up I jumped at the chance to join them. At Sun, the long hours continued into the early days of workstations and personal computers, and I have enjoyed participating in the creation of advanced microprocessor technologies and Internet technologies such as Java and Jini.

From all this, I trust it is clear that I am not a Luddite. I have always, rather, had a strong belief in the value of the scientific search for truth and in the ability of great engineering to bring material progress. The Industrial Revolution has immeasurably improved everyone's life over the last couple hundred years, and I always expected my career to involve the building of worthwhile solutions to real problems, one problem at a time.

I have not been disappointed. My work has had more impact than I had ever hoped for and has been more widely used than I could have reasonably expected. I have spent the last 20 years still trying to figure out how to make computers as reliable as I want them to be (they are not nearly there yet) and how to make them simple to use (a goal that has met with even less relative success). Despite some progress, the problems that remain seem even more daunting.

But while I was aware of the moral dilemmas surrounding technology's consequences in fields like weapons research, I did not expect that I would confront such issues in my own field, or at least not so soon.

Perhaps it is always hard to see the bigger impact while you are in the vortex of a change. Failing to understand the consequences of our inventions while we are in the rapture of discovery and innovation seems to be a common fault of scientists and technologists; we have long been driven by the overarching desire to know that is the nature of science's quest, not stopping to notice that the progress to newer and more powerful technologies can take on a life of its own.

I have long realized that the big advances in information technology come not from the work of computer scientists, computer architects, or electrical engineers, but from that of physical scientists. The physicists Stephen Wolfram and Brosl Hasslacher introduced me, in the early 1980s, to chaos theory and nonlinear systems. In the 1990s, I learned about complex systems from conversations with Danny Hillis, the biologist Stuart Kauffman, the Nobel-laureate physicist Murray Gell-Mann, and others. Most recently, Hasslacher and the electrical engineer and device physicist Mark Reed have been giving me insight into the incredible possibilities of molecular electronics.

In my own work, as codesigner of three microprocessor architectures—SPARC, picoJava, and MAJC—and as the designer of several implementations thereof, I've been afforded a deep and firsthand acquaintance with Moore's law. For decades, Moore's law has correctly predicted the exponential rate of improvement of semiconductor technology. Until last year I believed that the rate of advances predicted by Moore's law might continue only until roughly 2010, when some physical limits would begin to be reached. It was not obvious to me that a new technology would arrive in time to keep performance advancing smoothly.

But because of the recent rapid and radical progress in molecular electronics—where individual atoms and molecules replace lithographically drawn transistors—and related nanoscale technologies, we should be able to meet or exceed the Moore's law rate of progress for another 30 years. By 2030, we are likely to be able to build machines, in quantity, a million times as powerful as the personal computers of today—sufficient to implement the dreams of Kurzweil and Moravec.

As this enormous computing power is combined with the manipulative advances of the physical sciences and the new, deep understandings in genetics, enormous transformative power is being unleashed. These combinations open up the opportunity to completely redesign the world, for better or worse: The replicating and evolving processes that have been confined to the natural world are about to become realms of human endeavor.

In designing software and microprocessors, I have never had the feeling that I was designing an intelligent machine. The software and hardware is so fragile and the capabilities of the machine to "think" so clearly absent that, even as a possibility, this has always seemed very far in the future.

But now, with the prospect of human-level computing power in about 30 years, a new idea suggests itself: that I may be working to create tools which will enable the construction of the technology that may replace our species. How do I feel about this? Very uncomfortable. Having struggled my entire career to build reliable software systems, it seems to me more than likely that this future will not work out as well as some people may imagine. My personal experience suggests we tend to overestimate our design abilities.

Given the incredible power of these new technologies, shouldn't we be asking how we can best coexist with them? And if our own extinction is a likely, or even possible, outcome of our technological development, shouldn't we proceed with great caution?

The dream of robotics is, first, that intelligent machines can do our work for us, allowing us lives of leisure, restoring us to Eden. Yet in his history of such ideas, *Darwin Among the Machines,* George Dyson warns: "In the game of life and evolution there are three players at the table: human beings, nature, and machines. I am firmly on the side of nature. But nature, I suspect, is on the side of the machines." As we have seen, Moravec agrees, believing we may well not survive the encounter with the superior robot species.

How soon could such an intelligent robot be built? The coming advances in computing power seem to make it possible by 2030. And once an intelligent robot exists, it is only a small step to a robot species—to an intelligent robot that can make evolved copies of itself.

A second dream of robotics is that we will gradually replace ourselves with our robotic technology, achieving near immortality by downloading our consciousnesses; it is this process that Danny Hillis thinks we will gradually get used to and that Ray Kurzweil elegantly details in *The Age of Spiritual Machines.* (We are beginning to see intimations of this in the implantation of computer devices into the human body, as illustrated on the cover of *Wired* 8.02.)

But if we are downloaded into our technology, what are the chances that we will thereafter be ourselves or even human? It seems to me far more likely that a robotic existence would not be like a human one in any sense that we understand, that the robots would in no sense be our children, that on this path our humanity may well be lost.

Genetic engineering promises to revolutionize agriculture by increasing crop yields while reducing the use of pesticides; to create tens of thousands of novel species of bacteria, plants, viruses, and animals; to replace reproduction, or supplement it, with cloning; to create cures for many diseases, increasing our life span and our quality of life; and much, much more. We now know with certainty that these profound changes in the biological sciences are imminent and will challenge all our notions of what life is.

Technologies such as human cloning have in particular raised our awareness of the profound ethical and moral issues we face. If, for example, we were to reengineer ourselves into several separate and unequal species using the power of genetic engineering, then we would threaten the notion of equality that is the very cornerstone of our democracy.

Given the incredible power of genetic engineering, it's no surprise that there are significant safety issues in its use. My friend Amory Lovins recently cowrote, along with Hunter Lovins, an editorial that provides an ecological view of some of these dangers. Among their concerns: that "the new botany aligns the development of plants with their economic, not evolutionary, success." (See "A Tale of Two Botanies," page 247.) Amory's long career has been focused on energy and resource efficiency by taking a whole-system view of human-made systems; such a whole-system view often finds simple, smart solutions to otherwise seemingly difficult problems, and is usefully applied here as well.

After reading the Lovins' editorial, I saw an op-ed by Gregg Easterbrook in *The New York Times* (November 19, 1999) about genetically engineered crops, under the headline: "Food for the Future: Someday, rice will have built-in vitamin A. Unless the Luddites win."

Are Amory and Hunter Lovins Luddites? Certainly not. I believe we all would agree that golden rice, with its built-in vitamin A, is probably a good thing, if developed with proper care and respect for the likely dangers in moving genes across species boundaries.

Awareness of the dangers inherent in genetic engineering is beginning to grow, as reflected in the Lovins' editorial. The general public is aware of, and uneasy about, genetically modified foods, and seems to be rejecting the notion that such foods should be permitted to be unlabeled.

But genetic engineering technology is already very far along. As the Lovins note, the USDA has already approved about 50 genetically engineered crops for unlimited release; more than half of the world's soybeans and a third of its corn now contain genes spliced in from other forms of life.

While there are many important issues here, my own major concern with genetic engineering is narrower: that it gives the power—whether militarily, accidentally, or in a deliberate terrorist act—to create a White Plague.

The many wonders of nanotechnology were first imagined by the Nobel-laureate physicist Richard Feynman in a speech he gave in 1959, subsequently published under the title "There's Plenty of Room at the Bottom." The book that made a big impression on me, in the mid-'80s, was Eric Drexler's *Engines of Creation,* in which he described beautifully how manipulation of matter at the atomic level could create a utopian future of abundance, where just about everything could be made cheaply, and almost any imaginable disease or physical problem could be solved using nanotechnology and artificial intelligences.

A subsequent book, *Unbounding the Future: The Nanotechnology Revolution,* which Drexler cowrote, imagines some of the changes that might take place in a world where we had molecular-level "assemblers." Assemblers could make possible incredibly low-cost solar power, cures for cancer and the common cold by augmentation of the human immune system, essentially complete cleanup of the environment, incredibly inexpensive pocket supercomputers—in fact, any product would be manufacturable by assemblers at a cost no greater than that of wood—spaceflight more accessible than transoceanic travel today, and restoration of extinct species.

I remember feeling good about nanotechnology after reading *Engines of Creation.* As a technologist, it gave me a sense of calm—that is, nanotechnology showed us that incredible progress was possible, and indeed perhaps inevitable. If nanotechnology was our future, then I didn't feel pressed to solve so many problems in the present. I would get to Drexler's utopian future in due time; I might as well enjoy life more in the here and now. It didn't make sense, given his vision, to stay up all night, all the time.

Drexler's vision also led to a lot of good fun. I would occasionally get to describe the wonders of nanotechnology to others who had not heard of it. After teasing them with all the things Drexler described I would give a homework assignment of my own: "Use nanotechnology to create a vampire; for extra credit create an antidote."

With these wonders came clear dangers, of which I was acutely aware. As I said at a nanotechnology conference in 1989, "We can't simply do our science and not worry about these ethical issues."[5] But my subsequent conversations with physicists convinced me that nanotechnology might not even work—or, at least, it wouldn't work anytime soon. Shortly thereafter I moved to Colorado, to a skunk works I had set up, and the focus of my work shifted to software for the Internet, specifically on ideas that became Java and Jini.

Then, last summer, Brosl Hasslacher told me that nanoscale molecular electronics was now practical. This was new news, at least to me, and I think to many people—and it radically changed my opinion about nanotechnology. It sent me back to *Engines of Creation*. Rereading Drexler's work after more than 10 years, I was dismayed to realize how little I had remembered of its lengthy section called "Dangers and Hopes," including a discussion of how nanotechnologies can become "engines of destruction." Indeed, in my rereading of this cautionary material today, I am struck by how naive some of Drexler's safeguard proposals seem, and how much greater I judge the dangers to be now than even he seemed to then. (Having anticipated and described many technical and political problems with nanotechnology, Drexler started the Foresight Institute in the late 1980s "to help prepare society for anticipated advanced technologies"—most important, nanotechnology.)

The enabling breakthrough to assemblers seems quite likely within the next 20 years. Molecular electronics—the new subfield of nanotechnology where individual molecules are circuit elements—should mature quickly and become enormously lucrative within this decade, causing a large incremental investment in all nanotechnologies.

Unfortunately, as with nuclear technology, it is far easier to create destructive uses for nanotechnology than constructive ones. Nanotechnology has clear military and terrorist uses, and you need not be suicidal to release a massively destructive nanotechnological device—such devices can be built to be selectively destructive, affecting, for example, only a certain geographical area or a group of people who are genetically distinct.

An immediate consequence of the Faustian bargain in obtaining the great power of nanotechnology is that we run a grave risk—the risk that we might destroy the biosphere on which all life depends.

As Drexler explained:

> "Plants" with "leaves" no more efficient than today's solar cells could out-compete real plants, crowding the biosphere with an inedible foliage. Tough omnivorous "bacteria" could out-compete real bacteria: They could spread like blowing pollen, replicate swiftly, and reduce the biosphere to dust in a matter of days. Dangerous replicators could easily be too tough, small, and rapidly spreading to stop—at least if we make no preparation. We have trouble enough controlling viruses and fruit flies.

> Among the cognoscenti of nanotechnology, this threat has become known as the "gray goo problem." Though masses of uncontrolled replicators need not be gray or gooey, the term "gray goo" emphasizes that replicators able to obliterate life might be less inspiring than a single species of crabgrass. They might be superior in an evolutionary sense, but this need not make them valuable.
>
> The gray goo threat makes one thing perfectly clear: We cannot afford certain kinds of accidents with replicating assemblers.

Gray goo would surely be a depressing ending to our human adventure on Earth, far worse than mere fire or ice, and one that could stem from a simple laboratory accident.[6] Oops.

It is most of all the power of destructive self-replication in genetics, nanotechnology, and robotics (GNR) that should give us pause. Self-replication is the modus operandi of genetic engineering, which uses the machinery of the cell to replicate its designs, and the prime danger underlying gray goo in nanotechnology. Stories of run-amok robots like the Borg, replicating or mutating to escape from the ethical constraints imposed on them by their creators, are well established in our science fiction books and movies. It is even possible that self-replication may be more fundamental than we thought, and hence harder—or even impossible—to control. A recent article by Stuart Kauffman in *Nature* titled "Self-Replication: Even Peptides Do It" discusses the discovery that a 32-amino-acid peptide can "autocatalyse its own synthesis." We don't know how widespread this ability is, but Kauffman notes that it may hint at "a route to self-reproducing molecular systems on a basis far wider than Watson-Crick base-pairing."[7]

In truth, we have had in hand for years clear warnings of the dangers inherent in widespread knowledge of GNR technologies—of the possibility of knowledge alone enabling mass destruction. But these warnings haven't been widely publicized; the public discussions have been clearly inadequate. There is no profit in publicizing the dangers.

The nuclear, biological, and chemical (NBC) technologies used in 20th-century weapons of mass destruction were and are largely military, developed in government laboratories. In sharp contrast, the 21st-century GNR technologies have clear commercial uses and are being developed almost exclusively by corporate enterprises. In this age of triumphant commercialism, technology—with science as its handmaiden—is delivering a series of almost magical inventions that are the most phenomenally lucrative ever seen. We are aggressively pursuing the promises of these new technologies within the now-unchallenged system of global capitalism and its manifold financial incentives and competitive pressures.

> This is the first moment in the history of our planet when any species, by its own voluntary actions, has become a danger to itself—as well as to vast numbers of others.
>
> It might be a familiar progression, transpiring on many worlds—a planet, newly formed, placidly revolves around its star; life slowly forms; a kaleidoscopic procession of creatures evolves; intelligence emerges which, at least up to a point, confers enormous survival value; and then technology is invented. It dawns on them that there are such things as laws of Nature, that these laws can be revealed by experiment, and that knowledge of these laws can be made both to save and to take lives, both on unprecedented scales. Science, they recognize, grants immense powers. In a flash, they create world-altering contrivances. Some planetary civilizations see their way through, place limits on what may and what must not be done, and safely pass through the time of perils. Others, not so lucky or so prudent, perish.

That is Carl Sagan, writing in 1994, in *Pale Blue Dot,* a book describing his vision of the human future in space. I am only now realizing how deep his insight was, and how sorely I miss, and will miss, his voice. For all its eloquence, Sagan's contribution was not least that of simple common sense—an attribute that, along with humility, many of the leading advocates of the 21st-century technologies seem to lack.

I remember from my childhood that my grandmother was strongly against the overuse of antibiotics. She had worked since before the first World War as a nurse and had a commonsense attitude that taking antibiotics, unless they were absolutely necessary, was bad for you.

It is not that she was an enemy of progress. She saw much progress in an almost 70-year nursing career; my grandfather, a diabetic, benefited greatly from the improved treatments that became available in his lifetime. But she, like many levelheaded people, would probably think it greatly arrogant for us, now, to be designing a robotic "replacement species," when we obviously have so much trouble making relatively simple things work, and so much trouble managing—or even understanding—ourselves.

I realize now that she had an awareness of the nature of the order of life, and of the necessity of living with and respecting that order. With this respect comes a necessary humility that we, with our early-21st-century *chutzpah,* lack at our peril. The commonsense view, grounded in this respect, is often right, in advance of the scientific evidence. The clear fragility and inefficiencies of the human-made systems we have built should give us all pause; the fragility of the systems I have worked on certainly humbles me.

We should have learned a lesson from the making of the first atomic bomb and the resulting arms race. We didn't do well then, and the parallels to our current situation are troubling.

The effort to build the first atomic bomb was led by the brilliant physicist J. Robert Oppenheimer. Oppenheimer was not naturally interested in politics but became painfully aware of what he perceived as the grave threat to Western civilization from the Third Reich, a threat surely grave because of the possibility that Hitler might obtain nuclear weapons. Energized by this concern, he brought his strong intellect, passion for physics, and charismatic leadership skills to Los Alamos and led a rapid and successful effort by an incredible collection of great minds to quickly invent the bomb.

What is striking is how this effort continued so naturally after the initial impetus was removed. In a meeting shortly after V-E Day with some physicists who felt that perhaps the effort should stop, Oppenheimer argued to continue. His stated reason seems a bit strange: not because of the fear of large casualties from an invasion of Japan, but because the United Nations, which was soon to be formed, should have foreknowledge of atomic weapons. A more likely reason the project continued is the momentum that had built up—the first atomic test, Trinity, was nearly at hand.

We know that in preparing this first atomic test the physicists proceeded despite a large number of possible dangers. They were initially worried, based on a calculation by Edward Teller, that an atomic explosion might set fire to the atmosphere. A revised calculation reduced the danger of destroying the world to a three-in-a-million chance. (Teller says he was later able to dismiss the prospect of atmospheric ignition entirely.) Oppenheimer, though, was sufficiently concerned about the result of Trinity that he arranged for a possible evacuation of the southwest part of the state of New Mexico. And, of course, there was the clear danger of starting a nuclear arms race.

Within a month of that first, successful test, two atomic bombs destroyed Hiroshima and Nagasaki. Some scientists had suggested that the bomb simply be demonstrated, rather than dropped on Japanese cities—saying that this would greatly improve the chances for arms control after the war—but to no avail. With the tragedy of Pearl Harbor still fresh in Americans' minds, it would have been very difficult for President Truman to order a demonstration of the weapons rather than use them as he did—the desire to quickly end the war and save the lives that would have been lost in any invasion of Japan was very strong. Yet the overriding truth was probably very simple: As the physicist Freeman Dyson later said, "The reason that it was dropped was just that nobody had the courage or the foresight to say no."

It's important to realize how shocked the physicists were in the aftermath of the bombing of Hiroshima, on August 6, 1945. They describe a series of waves of emotion: first, a sense of fulfillment that the bomb worked, then horror at all the people that had been killed, and then a convincing feeling that on no account should another bomb be dropped. Yet of course another bomb was dropped, on Nagasaki, only three days after the bombing of Hiroshima.

In November 1945, three months after the atomic bombings, Oppenheimer stood firmly behind the scientific attitude, saying, "It is not possible to be a scientist unless you believe that the knowledge of the world, and the power which this gives, is a thing which is of intrinsic value to humanity, and that you are using it to help in the spread of knowledge and are willing to take the consequences."

Oppenheimer went on to work, with others, on the Acheson-Lilienthal report, which, as Richard Rhodes says in his recent book *Visions of Technology,* "found a way to prevent a clandestine nuclear arms race without resorting to armed world government"; their suggestion was a form of relinquishment of nuclear weapons work by nation-states to an international agency.

This proposal led to the Baruch Plan, which was submitted to the United Nations in June 1946 but never adopted (perhaps because, as Rhodes suggests, Bernard Baruch had "insisted on burdening the plan with conventional sanctions," thereby inevitably dooming it, even though it would "almost certainly have been rejected by Stalinist Russia anyway"). Other efforts to promote sensible steps toward internationalizing nuclear power to prevent an arms race ran afoul either of

U.S. politics and internal distrust, or distrust by the Soviets. The opportunity to avoid the arms race was lost, and very quickly.

Two years later, in 1948, Oppenheimer seemed to have reached another stage in his thinking, saying, "In some sort of crude sense which no vulgarity, no humor, no overstatement can quite extinguish, the physicists have known sin; and this is a knowledge they cannot lose."

In 1949, the Soviets exploded an atom bomb. By 1955, both the U.S. and the Soviet Union had tested hydrogen bombs suitable for delivery by aircraft. And so the nuclear arms race began.

Nearly 20 years ago, in the documentary *The Day After Trinity,* Freeman Dyson summarized the scientific attitudes that brought us to the nuclear precipice:

> I have felt it myself. The glitter of nuclear weapons. It is irresistible if you come to them as a scientist. To feel it's there in your hands, to release this energy that fuels the stars, to let it do your bidding. To perform these miracles, to lift a million tons of rock into the sky. It is something that gives people an illusion of illimitable power, and it is, in some ways, responsible for all our troubles—this, what you might call technical arrogance, that overcomes people when they see what they can do with their minds.[8]

Now, as then, we are creators of new technologies and stars of the imagined future, driven—this time by great financial rewards and global competition—despite the clear dangers, hardly evaluating what it may be like to try to live in a world that is the realistic outcome of what we are creating and imagining.

In 1947, *The Bulletin of the Atomic Scientists* began putting a Doomsday Clock on its cover. For more than 50 years, it has shown an estimate of the relative nuclear danger we have faced, reflecting the changing international conditions. The hands on the clock have moved 15 times and today, standing at nine minutes to midnight, reflect continuing and real danger from nuclear weapons. The recent addition of India and Pakistan to the list of nuclear powers has increased the threat of failure of the nonproliferation goal, and this danger was reflected by moving the hands closer to midnight in 1998.

In our time, how much danger do we face, not just from nuclear weapons, but from all of these technologies? How high are the extinction risks?

The philosopher John Leslie has studied this question and concluded that the risk of human extinction is at least 30 percent,[9] while Ray Kurzweil believes we have "a better than even chance of making it through," with the caveat that he has "always been accused of being an optimist." Not only are these estimates not encouraging, but they do not include the probability of many horrid outcomes that lie short of extinction.

Faced with such assessments, some serious people are already suggesting that we simply move beyond Earth as quickly as possible. We would colonize the galaxy using von Neumann probes, which hop from star system to star system, replicating as they go. This step will almost certainly be necessary 5 billion years from now (or sooner if our solar system is disastrously impacted by the impending collision of our galaxy with the Andromeda galaxy within the next 3 billion years), but if we take Kurzweil and Moravec at their word it might be necessary by the middle of this century.

What are the moral implications here? If we must move beyond Earth this quickly in order for the species to survive, who accepts the responsibility for the fate of those (most of us, after all) who are left behind? And even if we scatter to the stars, isn't it likely that we may take our problems with us or find, later, that they have followed us? The fate of our species on Earth and our fate in the galaxy seem inextricably linked.

Another idea is to erect a series of shields to defend against each of the dangerous technologies. The Strategic Defense Initiative, proposed by the Reagan administration, was an attempt to design such a shield against the threat of a nuclear attack from the Soviet Union. But as Arthur C. Clarke, who was privy to discussions about the project, observed:

> Though it might be possible, at vast expense, to construct local defense systems that would 'only' let through a few percent of ballistic missiles, the much touted idea of a national umbrella was nonsense. Luis Alvarez, perhaps the greatest experimental physicist of this century, remarked to me that the advocates of such schemes were 'very bright guys with no common sense.'

Clarke continued:

> Looking into my often cloudy crystal ball, I suspect that a total defense might indeed be possible in a century or so. But the technology involved would produce, as a by-product, weapons so terrible that no one would bother with anything as primitive as ballistic missiles.[10]

In *Engines of Creation,* Eric Drexler proposed that we build an active nanotechnological shield—a form of immune system for the biosphere—to defend against dangerous replicators of all kinds that might escape from laboratories or otherwise be maliciously created. But the shield he proposed would itself be extremely dangerous—nothing could prevent it from developing autoimmune problems and attacking the biosphere itself.[11]

Similar difficulties apply to the construction of shields against robotics and genetic engineering. These technologies are too powerful to be shielded against in the time frame of interest; even if it were possible to implement defensive shields, the side effects of their development would be at least as dangerous as the technologies we are trying to protect against.

These possibilities are all thus either undesirable or unachievable or both. The only realistic alternative I see is relinquishment: to limit development of the technologies that are too dangerous, by limiting our pursuit of certain kinds of knowledge.

Yes, I know, knowledge is good, as is the search for new truths. We have been seeking knowledge since ancient times. Aristotle opened his *Metaphysics* with the simple statement: "All men by nature desire to know." We have, as a bedrock value in our society, long agreed on the value of open access to information, and recognize the problems that arise with attempts to restrict access to and development of knowledge. In recent times, we have come to revere scientific knowledge.

But despite the strong historical precedents, if open access to and unlimited development of knowledge henceforth puts us all in clear danger of extinction, then common sense demands that we reexamine even these basic, long-held beliefs.

It was Nietzsche who warned us, at the end of the 19th century, not only that God is dead but that "faith in science, which after all exists undeniably, cannot owe its origin to a calculus of utility; it must have originated in spite of the fact that the disutility and dangerousness of the 'will to truth,' of 'truth at any price' is proved to it constantly." It is this further danger that we now fully face—the consequences of our truth-seeking. The truth that science seeks can certainly be considered a dangerous substitute for God if it is likely to lead to our extinction.

If we could agree, as a species, what we wanted, where we were headed, and why, then we would make our future much less dangerous—then we might understand what we can and should relinquish. Otherwise, we can easily imagine an arms race developing over GNR technologies, as it did with the NBC technologies in the 20th century. This is perhaps the greatest risk, for once such a race begins, it's very hard to end it. This time—unlike during the Manhattan Project—we aren't in a war, facing an implacable enemy that is threatening our civilization; we are driven, instead, by our habits, our desires, our economic system, and our competitive need to know.

I believe that we all wish our course could be determined by our collective values, ethics, and morals. If we had gained more collective wisdom over the past few thousand years, then a dialogue to this end would be more practical, and the incredible powers we are about to unleash would not be nearly so troubling.

One would think we might be driven to such a dialogue by our instinct for self-preservation. Individuals clearly have this desire, yet as a species our behavior seems to be not in our favor. In dealing with the nuclear threat, we often spoke dishonestly to ourselves and to each other, thereby greatly increasing the risks. Whether this was politically motivated, or because we chose not to think ahead, or because when faced with such grave threats we acted irrationally out of fear, I do not know, but it does not bode well.

The new Pandora's boxes of genetics, nanotechnology, and robotics are almost open, yet we seem hardly to have noticed. Ideas can't be put back in a box; unlike uranium or plutonium, they don't need to be mined and refined, and they can be freely copied. Once they are out, they are out. Churchill remarked, in a famous left-handed compliment, that the American people and their leaders "invariably do the right thing, after they have examined every other alternative." In this case, however, we must act more presciently, as to do the right thing only at last may be to lose the chance to do it at all.

As Thoreau said, "We do not ride on the railroad; it rides upon us"; and this is what we must fight, in our time. The question is, indeed, Which is to be master? Will we survive our technologies?

We are being propelled into this new century with no plan, no control, no brakes. Have we already gone too far down the path to alter course? I don't believe so, but we aren't trying yet, and the last chance to assert control—the fail-safe point—is rapidly approaching. We have our first pet robots, as well as commercially available genetic engineering techniques, and our nanoscale techniques are advancing rapidly. While the development of these technologies proceeds through a number of steps, it isn't necessarily the case—as happened in the Manhattan Project and the Trinity test—that the last step in proving a technology is large and hard. The breakthrough to wild self-replication in robotics, genetic engineering, or nanotechnology could come suddenly, reprising the surprise we felt when we learned of the cloning of a mammal.

And yet I believe we do have a strong and solid basis for hope. Our attempts to deal with weapons of mass destruction in the last century provide a shining example of relinquishment for us to consider: the unilateral U.S. abandonment, without preconditions, of the development of biological weapons. This relinquishment stemmed from the realization that while it would take an enormous effort to create these terrible weapons, they could from then on easily be duplicated and fall into the hands of rogue nations or terrorist groups.

The clear conclusion was that we would create additional threats to ourselves by pursuing these weapons, and that we would be more secure if we did not pursue them. We have embodied our relinquishment of biological and chemical weapons in the 1972 Biological Weapons Convention (BWC) and the 1993 Chemical Weapons Convention (CWC).[12]

As for the continuing sizable threat from nuclear weapons, which we have lived with now for more than 50 years, the U.S. Senate's recent rejection of the Comprehensive Test Ban Treaty makes it clear relinquishing nuclear weapons will not be politically easy. But we have a unique opportunity, with the end of the Cold War, to avert a multipolar arms race. Building on the BWC and CWC relinquishments, successful abolition of nuclear weapons could help us build toward a habit of relinquishing dangerous technologies. (Actually, by getting rid of all but 100 nuclear weapons worldwide—roughly the total destructive power of World War II and a considerably easier task—we could eliminate this extinction threat.[13])

Verifying relinquishment will be a difficult problem, but not an unsolvable one. We are fortunate to have already done a lot of relevant work in the context of the BWC and other treaties. Our major task will be to apply this to technologies that are naturally much more commercial than military. The substantial need here is for transparency, as difficulty of verification is directly proportional to the difficulty of distinguishing relinquished from legitimate activities.

I frankly believe that the situation in 1945 was simpler than the one we now face: The nuclear technologies were reasonably separable into commercial and military uses, and monitoring was aided by the nature of atomic tests and the ease with which radioactivity could be measured. Research on military applications could be performed at national laboratories such as Los Alamos, with the results kept secret as long as possible.

The GNR technologies do not divide clearly into commercial and military uses; given their potential in the market, it's hard to imagine pursuing them only in national laboratories. With their widespread commercial pursuit, enforcing relinquishment will require a verification regime similar to that for biological weapons, but on an unprecedented scale. This, inevitably, will raise tensions between our individual privacy and desire for proprietary information, and the need for verification to protect us all. We will undoubtedly encounter strong resistance to this loss of privacy and freedom of action.

Verifying the relinquishment of certain GNR technologies will have to occur in cyberspace as well as at physical facilities. The critical issue will be to make the necessary transparency acceptable in a world of proprietary information, presumably by providing new forms of protection for intellectual property.

Verifying compliance will also require that scientists and engineers adopt a strong code of ethical conduct, resembling the Hippocratic oath, and that they have the courage to whistleblow as necessary, even at high personal cost. This would answer the call—50 years after Hiroshima—by the Nobel laureate Hans Bethe, one of the most senior of the surviving members of the Manhattan Project, that all scientists "cease and desist from work creating, developing, improving, and manufacturing nuclear weapons and other weapons of potential mass destruction."[14] In the 21st century, this requires vigilance and personal responsibility by those who would work on both NBC and GNR technologies to avoid implementing weapons of mass destruction and knowledge-enabled mass destruction.

Thoreau also said that we will be "rich in proportion to the number of things which we can afford to let alone." We each seek to be happy, but it would seem worthwhile to question whether we need to take such a high risk of total destruction to gain yet more knowledge and yet more things; common sense says that there is a limit to our material needs—and that certain knowledge is too dangerous and is best forgone.

Neither should we pursue near immortality without considering the costs, without considering the commensurate increase in the risk of extinction. Immortality, while perhaps the original, is certainly not the only possible utopian dream.

I recently had the good fortune to meet the distinguished author and scholar Jacques Attali, whose book *Lignes d'horizons* (*Millennium,* in the English translation) helped inspire the Java and Jini approach to the coming age of pervasive computing, as previously described in this magazine. In his new book *Fraternités,* Attali describes how our dreams of utopia have changed over time:

> At the dawn of societies, men saw their passage on Earth as nothing more than a labyrinth of pain, at the end of which stood a door leading, via their death, to the company of gods and to *Eternity.* With the Hebrews and then the Greeks, some men dared free themselves from theological demands and dream of an ideal City where *Liberty* would flourish. Others, noting the evolution of the market society, understood that the liberty of some would entail the alienation of others, and they sought *Equality.*

Jacques helped me understand how these three different utopian goals exist in tension in our society today. He goes on to describe a fourth utopia, *Fraternity,* whose foundation is altruism. *Fraternity* alone associates individual happiness with the happiness of others, affording the promise of self-sustainment.

This crystallized for me my problem with Kurzweil's dream. A technological approach to *Eternity*—near immortality through robotics—may not be the most desirable utopia, and its pursuit brings clear dangers. Maybe we should rethink our utopian choices.

Where can we look for a new ethical basis to set our course? I have found the ideas in the book *Ethics for the New Millennium,* by the Dalai Lama, to be very helpful. As is perhaps well known but little heeded, the Dalai Lama argues that the most important thing is for us to conduct our lives with love and compassion for others, and that our societies need to develop a stronger notion of universal responsibility and of our interdependency; he proposes a standard of positive ethical conduct for individuals and societies that seems consonant with Attali's *Fraternity* utopia.

The Dalai Lama further argues that we must understand what it is that makes people happy, and acknowledge the strong evidence that neither material progress nor the pursuit of the power of knowledge is the key—that there are limits to what science and the scientific pursuit alone can do.

Our Western notion of happiness seems to come from the Greeks, who defined it as "the exercise of vital powers along lines of excellence in a life affording them scope."[15]

Clearly, we need to find meaningful challenges and sufficient scope in our lives if we are to be happy in whatever is to come. But I believe we must find alternative outlets for our creative forces, beyond the culture of perpetual economic growth; this growth has largely been a blessing for several hundred years, but it has not brought us unalloyed happiness, and we must now choose between the pursuit of unrestricted and undirected growth through science and technology and the clear accompanying dangers.

It is now more than a year since my first encounter with Ray Kurzweil and John Searle. I see around me cause for hope in the voices for caution and relinquishment and in those people I have discovered who are as concerned as I am about our current predicament. I feel, too, a deepened sense of personal responsibility—not for the work I have already done, but for the work that I might yet do, at the confluence of the sciences.

But many other people who know about the dangers still seem strangely silent. When pressed, they trot out the "this is nothing new" riposte—as if awareness of what could happen is response enough. They tell me, There are universities filled with bioethicists who study this stuff all day long. They say, All this has been written about before, and by experts. They complain, Your worries and your arguments are already old hat.

I don't know where these people hide their fear. As an architect of complex systems I enter this arena as a generalist. But should this diminish my concerns? I am aware of how much has been written about, talked about, and lectured about so authoritatively. But does this mean it has reached people? Does this mean we can discount the dangers before us?

Knowing is not a rationale for not acting. Can we doubt that knowledge has become a weapon we wield against ourselves?

The experiences of the atomic scientists clearly show the need to take personal responsibility, the danger that things will move too fast, and the way in which a process can take on a life of its own. We can, as they did, create insurmountable problems in almost no time flat. We must do more thinking up front if we are not to be similarly surprised and shocked by the consequences of our inventions.

My continuing professional work is on improving the reliability of software. Software is a tool, and as a toolbuilder I must struggle with the uses to which the tools I make are put. I have always believed that making software more reliable, given its many uses, will make the world a safer and better place; if I were to come to believe the opposite, then I would be morally obligated to stop this work. I can now imagine such a day may come.

This all leaves me not angry but at least a bit melancholic. Henceforth, for me, progress will be somewhat bittersweet.

Do you remember the beautiful penultimate scene in *Manhattan* where Woody Allen is lying on his couch and talking into a tape recorder? He is writing a short story about people who are creating unnecessary, neurotic problems for themselves, because it keeps them from dealing with more unsolvable, terrifying problems about the universe.

He leads himself to the question, "Why is life worth living?" and to consider what makes it worthwhile for him: Groucho Marx, Willie Mays, the second movement of the Jupiter Symphony, Louis Armstrong's recording of "Potato Head Blues," Swedish movies, Flaubert's Sentimental Education, Marlon Brando, Frank Sinatra, the apples and pears by Cezanne, the crabs at Sam Wo's, and, finally, the showstopper: his love Tracy's face.

Each of us has our precious things, and as we care for them we locate the essence of our humanity. In the end, it is because of our great capacity for caring that I remain optimistic we will confront the dangerous issues now before us.

My immediate hope is to participate in a much larger discussion of the issues raised here, with people from many different backgrounds, in settings not predisposed to fear or favor technology for its own sake.

As a start, I have twice raised many of these issues at events sponsored by the Aspen Institute and have separately proposed that the American Academy of Arts and Sciences take them up as an extension of its work with the Pugwash Conferences. (These have been held since 1957 to discuss arms control, especially of nuclear weapons, and to formulate workable policies.)

It's unfortunate that the Pugwash meetings started only well after the nuclear genie was out of the bottle—roughly 15 years too late. We are also getting a belated start on seriously addressing the issues around 21st-century technologies—the prevention of knowledge-enabled mass destruction—and further delay seems unacceptable.

So I'm still searching; there are many more things to learn. Whether we are to succeed or fail, to survive or fall victim to these technologies, is not yet decided. I'm up late again—it's almost 6 a.m. I'm trying to imagine some better answers, to break the spell and free them from the stone.

Endnotes

1. The passage Kurzweil quotes is from Kaczynski's "Unabomber Manifesto," which was published jointly, under duress, by *The New York Times* and *The Washington Post* to attempt to bring his campaign of terror to an end. I agree with David Gelernter, who said about their decision:

   "It was a tough call for the newspapers. To say yes would be giving in to terrorism, and for all they knew he was lying anyway. On the other hand, to say yes might stop the killing. There was also a chance that someone would read the tract and get a hunch about the author; and that is exactly what happened. The suspect's brother read it, and it rang a bell."

   "I would have told them not to publish. I'm glad they didn't ask me. I guess." (*Drawing Life: Surviving the Unabomber.* Free Press, 1997: 120.)
2. Garrett, Laurie. *The Coming Plague: Newly Emerging Diseases in a World Out of Balance.* Penguin, 1994: 47-52, 414, 419, 452.
3. Isaac Asimov described what became the most famous view of ethical rules for robot behavior in his book *I, Robot* in 1950, in his Three Laws of Robotics: 1. A robot may not injure a human being, or, through inaction, allow a human being to come to harm. 2. A robot must obey the orders given it by human beings, except where such orders would conflict with the First Law. 3. A robot must protect its own existence, as long as such protection does not conflict with the First or Second Law.
4. Michelangelo wrote a sonnet that begins:

   > Non ha l' ottimo artista alcun concetto Ch' un marmo solo in se non circonscriva Col suo soverchio; e solo a quello arriva La man che ubbidisce all' intelleto.

   Stone translates this as:

   > The best of artists hath no thought to show which the rough stone in its superfluous shell doth not include; to break the marble spell is all the hand that serves the brain can do.

   Stone describes the process:

   > He was not working from his drawings or clay models; they had all been put away. He was carving from the images in his mind. His eyes and hands knew where every line, curve, mass must emerge, and at what depth in the heart of the stone to create the low relief.

   (*The Agony and the Ecstasy.* Doubleday, 1961: 6, 144.)
5. First Foresight Conference on Nanotechnology in October 1989, a talk titled The Future of Computation. Published in Crandall, B. C. and James Lewis, editors. *Nanotechnology: Research and Perspectives.* MIT Press, 1992: 269. See also www.foresight.org/Conferences/MNT01/Nano1.html.
6. In his 1963 novel *Cat's Cradle,* Kurt Vonnegut imagined a gray-goo-like accident where a form of ice called ice-nine, which becomes solid at a much higher temperature, freezes the oceans.

7. Kauffman, Stuart. "Self-replication: Even Peptides Do It." *Nature,* 382, August 8, 1996: 496. See www.santafe.edu/sfi/People/kauffman/sak-peptides.html.

8. Else, Jon. *The Day After Trinity: J. Robert Oppenheimer and The Atomic Bomb* (available at www.pyramiddirect.com).

9. This estimate is in Leslie's book *The End of the World: The Science and Ethics of Human Extinction,* where he notes that the probability of extinction is substantially higher if we accept Brandon Carter's Doomsday Argument, which is, briefly, that "we ought to have some reluctance to believe that we are very exceptionally early, for instance in the earliest 0.001 percent, among all humans who will ever have lived. This would be some reason for thinking that humankind will not survive for many more centuries, let alone colonize the galaxy. Carter's doomsday argument doesn't generate any risk estimates just by itself. It is an argument for revising the estimates which we generate when we consider various possible dangers." (Routledge, 1996: 1, 3, 145.)

10. Clarke, Arthur C. "Presidents, Experts, and Asteroids."*Science,* June 5, 1998. Reprinted as "Science and Society" in *Greetings, Carbon-Based Bipeds! Collected Essays,* 1934–1998. St. Martin's Press, 1999: 526.

11. And, as David Forrest suggests in his paper "Regulating Nanotechnology Development," available at www.foresight.org/NanoRev/Forrest1989.html, "If we used strict liability as an alternative to regulation it would be impossible for any developer to internalize the cost of the risk (destruction of the biosphere), so theoretically the activity of developing nanotechnology should never be undertaken." Forrest's analysis leaves us with only government regulation to protect us—not a comforting thought.

12. Meselson, Matthew. "The Problem of Biological Weapons." Presentation to the 1,818th Stated Meeting of the American Academy of Arts and Sciences, January 13, 1999. (minerva.amacad.org/archive/bulletin4.htm)

13. Doty, Paul. "The Forgotten Menace: Nuclear Weapons Stockpiles Still Represent the Biggest Threat to Civilization."*Nature,* 402, December 9, 1999: 583.

14. See also Hans Bethe's 1997 letter to President William Clinton, at www.fas.org/bethecr.htm.

15. Hamilton, Edith. *The Greek Way.* W. W. Norton & Co., 1942: 35.

4 A Response to Bill Joy and the Doom-and-Gloom Technofuturists

John Seely Brown and Paul Duguid

If you lived through the 1950s, you might remember President Eisenhower, orderly suburban housing tracts, backyard bomb shelters—and dreams of a nuclear power plant in every home. Plans for industrial nuclear generators had barely left the drawing board before futurists predicted that every house would have a miniature version. From there, technoenthusiasts predicted the end of power monopolies, the emergence of the "electronic cottage," the death of the city and the decline of the corporation.

Pessimists and luddites, of course, envisioned nuclear apocalypse. Each side waited for nirvana, or Armageddon, so it could triumphantly tell the other, "I told you so."

With "Why the Future Doesn't Need Us" in the April issue of *Wired*, Bill Joy invokes those years gone by. No luddite, Joy is an awe-inspiring technologist—as cofounder and chief scientist of Sun Microsystems, he coauthored, among other things, the Java programming language. So when his article describes a technological juggernaut thundering toward society—bringing with it mutant genes, molecular-level nanotechnology machines and superintelligent robots—all need to listen. Like the nuclear prognosticators, Joy can see the juggernaut clearly. What he can't see—which is precisely what makes his vision so scary—are any controls.

John Seely Brown is chief scientist of the Xerox Corporation, and director of the Xerox Palo Alto Research Center (PARC). Paul Duguid is a research specialist in the division of Social and Cultural Studies in Education at the University of California, Berkeley, and a consultant at the Xerox PARC.

But it doesn't follow that the juggernaut is uncontrollable. To understand why not, readers should note the publication in which this article appeared. For the better part of a decade, *Wired* has been a cheerleader for the digital age. Until now, *Wired* has rarely been a venue to which people have looked for a way to put a brake on innovation. Therefore its shift with Joy's article from cheering to warning marks an important and surprising moment in the digital zeitgeist.

In an effort to locate some controls, let's go back to the nuclear age. Innovation, the argument went back in the 1950s, would make nuclear power plants smaller and cheaper. They would enter mass production and quickly become available to all.

Even today the argument might appear inescapable until you notice what's missing: The tight focus of this vision makes it almost impossible to see forces other than technology at work. In the case of nuclear development, a host of forces worked to dismantle the dream of a peaceful atom, including the environmental movement, antinuclear protests, concerned scientists, worried neighbors of Chernobyl and Three Mile Island, government regulators and antiproliferation treaties. Cumulatively, these forces slowed the nuclear juggernaut to a crawl.

Similar social forces are at work on technologies today. But because the digerati, like technoenthusiasts before them, look to the future with technological tunnel vision, they too have trouble bringing other forces into view.

The Tunnel Ahead

In Joy's vision, as in the nuclear one, there's a recognizable tunnel vision that leaves people out of the picture and focuses on technology in splendid isolation. This vision leads not only to doom-and-gloom scenarios, but also to tunnel design: the design of "simple" technologies that are actually difficult to use.

To escape both trite scenarios and bad design, we have to widen our horizons and bring into view not only technological systems, but also social systems. Good designs look beyond the dazzling potential of the technology to social factors, such as the limited patience of most users.

Paying attention to the latter has, for example, allowed the PalmPilot and Nintendo Game Boy to sweep aside more complex rivals. Their elegant simplicity has made them readily usable. And their usability has in turn created an important social support system. The devices are so widely used that anyone having trouble with a Pilot or Game Boy rarely has to look far for advice from a more experienced user.

As this small example suggests, technological and social systems shape each other. The same is true on a larger scale. Technologies—such as gunpowder, the printing press, the railroad, the telegraph and the Internet—can shape society in profound ways. But, on the other hand, social systems—in the form of governments, the courts, formal and informal organizations, social movements, professional networks, local communities, market institutions and so forth—shape, moderate and redirect the raw power of technologies.

Given the crisp edges of technology and the fuzzy outlines of society, it certainly isn't easy to use these two worldviews simultaneously. But if you want to see where we are going, or design the means to get there, you need to grasp both.

This perspective allows a more sanguine look at Joy's central concerns: genetic engineering, nanotechnology and robotics. Undoubtedly, each deserves serious thought. But each should be viewed in the context of the social system in which it is inevitably embedded.

Genetic engineering presents the clearest example. Barely a year ago, the technology seemed to be an unstoppable force. Major chemical and agricultural interests were barreling down an open highway. In the past year, however, road conditions changed dramatically for the worse: Cargill faced Third World protests against its patents; Monsanto suspended research on sterile seeds; and champions of genetically modified foods, who once saw an unproblematic and lucrative future, are scurrying to counter consumer boycotts of their products.

Almost certainly, those who support genetic modification will have to look beyond the technology if they want to advance it. They need to address society directly—not just by putting labels on modified foods, but by educating people about the costs and the benefits of these new agricultural products. Having ignored social concerns, however, proponents have made the people they need to educate profoundly suspicious and hostile.

Nanotechnology offers a rather different example of how the future can frighten us. Because the technology involves engineering at a molecular level, both the promise and the threat seem immeasurable. But they are immeasurable for a good reason: The technology is still almost wholly on the drawing board.

Two of nanotechnology's main proponents, Ralph Merkle and Eric Drexler, worked with us at the Xerox Palo Alto Research Center in

Palo Alto, Calif. The two built powerful nano-CAD tools and then ran simulations of the resulting molecular-level designs. These experiments showed definitively that nano devices are theoretically feasible. No one, however, has laid out a route from lab-based simulation to practical systems in any detail.

In the absence of a plan, it's important to ask the right questions: Can nanotechnology fulfill its great potential in tasks ranging from data storage to pollution control, all without spiraling out of control? If the lesson of genetic engineering is any guide, planners would do well to consult and educate the public early on, even though useful nano systems are probably decades away.

Worries about robotics appear premature, as well. Internet "bots" that search, communicate and negotiate for their human masters may appear to behave like Homo sapiens, but in fact, bots are often quite inept at functions that humans do well—functions that call for judgment, discretion, initiative or tacit understanding. They are good (and useful) for those tasks that humans do poorly. So they are better thought of as complementary systems, not rivals to humanity. Although bots will undoubtedly get better at what they do, such development will not necessarily make them more human.

Are more conventional clanking robots—the villains of science fiction—any great threat to society? We doubt it. Xerox PARC research on self-aware, reconfigurable "polybots" has pushed the boundaries of what robots can do, pointing the way to "morphing robots" that are able to move and change shape.

Nonetheless, for all their cutting-edge agility, these robots are a long way from making good dance partners. The chattiness of *Star Wars'* C-3PO still lies well beyond real-world machines. Indeed, what talk robots or computers achieve, though it may appear similar, is quite different from human talk. Talking machines travel routes designed specifically to avoid the full complexities of human language.

Robots may seem intelligent, but such intelligence is profoundly hampered by their inability to learn in any significant way. (This failing has apparently led Toyota, after heavy investment in robotics, to consider replacing robots with humans on many production lines.) And without learning, simple common sense will lie beyond robots for a long time to come.

Indeed, despite years of startling advances and innumerable successes like the chess-playing Big Blue, computer science is still about as far as it ever was from building a machine with the learning abilities, linguistic competence, common sense or social skills of a 5-year-old child.

As with Internet bots, real-world robots will no doubt become increasingly useful. But they will probably also become increasingly frustrating to use as a result of tunnel design. In that regard, they may indeed seem antisocial, but not in the way of *Terminator*-like fantasies of robot armies that lay waste to human society.

Indeed, the thing that handicaps robots most is their lack of a social existence. For it is our social existence as humans that shapes how we speak, learn, think and develop common sense. All forms of artificial life (whether bugs or bots) will remain primarily a metaphor for—rather than a threat to—society, at least until they manage to enter a debate, sing in a choir, take a class, survive a committee meeting, join a union, pass a law, engineer a cartel or summon a constitutional convention.

These critical social mechanisms allow society to shape its future. It is through planned, collective action that society forestalls expected consequences (such as Y2K) and responds to unexpected events (such as epidemics).

The Failure of a "6-D" Vision

Why does the threat of a cunning, replicating robot society look so close from one perspective, yet so distant from another? The difference lies in the well-known tendency of futurologists to count "1, 2, 3 . . . a million." That is, once the first step on a path is taken, it's very easy to assume that all subsequent steps are trivial.

Several of the steps Joy asks us to take—the leap from genetic engineering to a "white plague"; from simulations to out-of-control nanotechnology; from replicating peptides to a "robot species"—are extremely large. And they are certainly not steps that will be taken without diversions, regulations or controls.

One of the lessons of Joy's article, then, is that the path to the future can look simple (and sometimes downright terrifying) if you look at it through what we call "6-D lenses." We coined this phrase having so often in our research hit up against upon such "de-" or "di-" words as demassification, decentralization, disintermediation, despacialization, disaggregation and demarketization in the canon of futurology.

If you take any one of these words in isolation, it's easy to follow their relentless logic to its evident conclusion. Because firms are getting smaller, for example, it's easy to assume that companies and other intermediaries are simply disintegrating into markets. And because communication is growing cheaper and more powerful, it's easy to believe in the "death of distance."

But things rarely work in such linear fashion. Other forces are often at work, such as those driving firms into larger and larger mergers to take advantage of social, rather than merely technological, networks. Similarly, even though communications technology has killed distance, people curiously can't stay away from the social hotbed of modern communications technology, Silicon Valley.

Importantly, these d-words indicate that the old ties that once bound communities, organizations and institutions are being picked apart by technologies. A simple, linear reading, then, suggests that these communities, organizations and institutions will now simply fall apart. A more complex reading, taking into account the multiple forces at work, offers another picture.

While many powerful national corporations have grown insignificant, some have transformed into more powerful transnational firms. While some forms of community may be dying, others, bolstered by technology, are growing stronger.

Technology and society are constantly forming and reforming new dynamic equilibriums with far-reaching implications. The challenge for futurology (and for all of us) is to see beyond the hype and past the oversimplifications to the full import of these new sociotechnical formations.

Two hundred years ago, Thomas Malthus, assuming that human society and agricultural technology developed on separate paths, predicted that society was growing so fast that it would starve itself to death, the so-called Malthusian trap.

A hundred years later, H.G. Wells similarly assumed that society and technology were developing independently. Like many people today, Wells saw the advance of technology outstripping the evolution of society, leading him to predict that technology's relentless juggernaut would unfeelingly crush society. Like Joy, both Malthus and Wells issued important warnings, alerting society to the dangers it faced. But by their actions, Malthus and Wells helped prevent the very future they were so certain would come about.

These self-*un*fulfilling prophecies failed to see that, once warned, society could galvanize itself into action. Of course, this social action in the face of threats showed that Malthus and Wells were most at fault in their initial assumption. Social and technological systems do not develop independently; the two evolve together in complex feedback loops, wherein each drives, restrains and accelerates change in the other. Malthus and Wells—and now Joy—are, indeed, critical parts of these complex loops. Each knew when and how to sound the alarm. But each thought little about how to respond to that alarm.

Once the social system is factored back into the equation like this, the road ahead becomes harder to navigate. Ultimately we should be grateful to Joy for saying, at the least, that there could be trouble ahead when so many of his fellow digerati will only tell us complacently that the road is clear.

5 Not by Reason Alone

Michael L. Dertouzos

In a recent *Wired* magazine article, Bill Joy argued that the consequences of research on robotics, genetic engineering and nanotechnology may lead to "knowledge-enabled mass destruction . . . hugely amplified by the power of self-replication." His medicine: "relinquishment . . . by limiting our pursuit of certain kinds of knowledge." I don't buy it.

What troubles me with this argument is the arrogant notion that human logic can anticipate the effects of intended or unintended acts, and the more arrogant notion that human reasoning can determine the course of the universe. Let me explain and offer some alternatives.

We are seldom able to assess where we are headed. In 1963, when we built time-shared computers, we did it to spread the cost of a $2 million processor among many users. In 1970, when DARPA pioneered the Arpanet, it did so to avoid buying expensive computers for its contractors, who were told to share their networked machines. Both efforts succeeded, not for these goals, but because they enabled people to share information. The Internet was launched to interconnect networks of computers—no one anticipated that its biggest application would be the Web. Radar was designed for war but ended up as a cornerstone of air transportation. Nuclear weapons research put nuclear medicine on the map. Thousands of innovations all share the same pattern—the early assessment is unrelated to the outcome.

So limited is our ability to assess consequences that it's not even helped by hindsight: On balance, are cars a good or bad thing for society? How about nuclear power, or nuclear medicine? We are unable to judge whether something we invented more than 50 years ago is good or bad for us today. Yet Joy wants us to make these judgments prospectively, to determine which technologies we should forgo!

Michael L. Dertouzos is director of the Laboratory for Computer Science and professor of computer science and electrical engineering at the Massachusetts Institute of Technology. *This article is reprinted with permission from* Technology Review, *September/October 2000, p. 26. Copyright 2000 MIT Press.*

Developments that today seem fearful may turn into mirages. Take the spiritual machines of Ray Kurzweil that concern Bill Joy. I have a lot of respect for Ray, and I welcome his ideas, as I do Bill's, however outlandish or controversial they may be. But we should draw a clear line between what is imagined and what is likely. To blur this line is tantamount to quackery. Just because chips and machines are getting faster doesn't mean they'll get smarter, let alone lead to self-replication. If you move your arms faster you won't get brighter. Despite fashionable hyperventilation about intelligent agents, today's computer systems are not intelligent in the normal sense of the word. Nor do we see on the research horizon the critical technologies that would lead them there. Should we stop computer science and AI research in the belief that intelligent machines someday will reproduce themselves and surpass us? I say no. We should wait to find out whether the potential dangers are supported by more than our imagination.

Since we can't see where we are headed, should we stop research altogether? This reminds me of a wise old airline employee to whom I was bragging that I had stopped flying with his company because of their lousy safety record. "Listen, sir," he said to me. "If your exit visa from this life is stamped 'death by aircraft,' even if you stay in your bed, the airplane will find you and crash upon you." At this, the dawn of the technology century, it is not fashionable to pay attention to forces outside reason. We should reconsider. All the more so, if we are under the illusion that we understand enough of our universe to successfully regulate its future course, as Joy suggests.

We shouldn't forget that what we do as human beings is part of nature. I am not advocating that we do as we please, on the grounds that it is natural, but rather that we hold nature—including our actions—in awe. As we fashion grand strategies to "regulate the ozone problem," or any other complex aspect of our world, we should be respectful of the unpredictable ways nature may react. And we should approach with equal respect the presumption that the natural human urge to probe our universe should be restricted.

I suggest we broaden our perspective to the fullness of our humanity, which besides reason includes feelings and beliefs. Sometimes, as we drive the car of scientific and technological progress, we'll veer because our reason says so. At other times we'll follow our feelings, or we'll be guided by faith. Most of the time, we'll steer with all three of these human forces guiding us in concert, as they have guided human actions for thousands of years. As we do so, we should stay vigilant, ready to stop, when danger is imminent, using our full humanity to make that determination. If we do so, our turning point will be very different from where it may seem today, based on early rational assessments...that have failed us so often. Let us have faith in ourselves, our fellow human beings and our universe. And let's keep in mind that our car is not the only moving thing out there.

6 Nanotechnology and Societal Transformation

Michael M. Crow and Daniel Sarewitz

Remaking the World

Technological innovation sustains a fundamental tension of civilization, the tension between humanity's quest for more control over nature and the future, and our equally strong desire for stability and predictability in the present. The original Luddites were not against technology per se. They were against losing their jobs, and so they smashed the power looms that had put them out of work. The change wrought by technological advance continually remakes society, and this transformational process is on the one hand central to the dynamic that is commonly labeled "progress," and yet on the other is a source of continual destabilization and dislocation as experienced by individuals, communities, institutions, nations, and cultures.

In the age of science and technology (S&T), the federal government has increasingly become the prime catalyst for scientific advance and technological innovation. At the same time, modern government is also continually responding to and managing the transformational power of science and technology. Yet there is little effort to understand the relation between these two critical activities—advancing knowledge and innovation, and responding to their impacts—or to link them in a way that can enhance the value and capability of each.

A single technological innovation can remake the world. When the metal stirrup finally migrated from Asia to western Europe in the 8th

---

Michael M. Crow is executive vice provost and professor of science policy at Columbia University. Daniel Sarewitz is managing director and senior research scholar at the Center for Science, Policy and Outcomes, a project of Columbia University. *This paper was presented at the National Science and Technology Council Workshop on Societal Implications of Nanoscience and Nanotechnology, Sept. 28–29, 2000.*

century, society was transformed to its very roots. For the first time, the energy of a galloping horse could be directly transmitted to the weapon held by the man in the saddle—a combat innovation of devastating impact. Because horses and tack were costly, they were possessed almost exclusively by landowners. Battlefield prowess and wealth were thus combined, and from this combination grew not just the traditions of a "warrior aristocracy" but the structure of European feudal society itself. Later, when the Anglo Saxon King Harold prepared to defend Britain against the invading Normans in 1066, he actually dispensed with his horse and ornamental wooden stirrups, choosing to lead his numerically superior forces on foot. The outnumbered Normans, however, boasted a strong, stirrup-equipped cavalry, and thus won the day—and the millennium.[1]

Such narrative has the ring of mythology, yet the experience of the industrialized world reinforces the knowledge that a new machine can help change everything. The invention of the cotton gin in the late 18th century allowed a vast expansion of cotton cultivation in the American south—and directly fueled a resurgence in the importation and use of slaves for plantation labor. One hundred and fifty years later, the mechanical cotton picker suddenly rendered obsolete the jobs of millions of African American share croppers, and catalyzed a 30-year migration of five million people out of the rural south and into the cities of the north. While the development of the mechanical cotton picker was no doubt inevitable, its proliferation was consciously accelerated by plantation owners who, fearing the rise of the civil rights movement, sought quickly to find a technological replacement for the existing system of exploitation labor upon which they were economically dependent.[2]

These examples point not only to the power of new technologies to transform society, but to the comprehensive interconnectedness of technological change and the complex social structure of society. The invention of the stirrup as a battlefield tool was in some very intricate way connected to the development and expansion of feudalism in Europe; the evolution of agricultural technology for a single cash crop is indissolubly bound to the ongoing struggle to overcome the U.S. legacy of slavery, segregation, and bigotry. More familiarly, a single class of technology—nuclear weapons—was a central determinant of geopolitical evolution after the end of World War II. Cars, television, air conditioning, and vaccinations have all stimulated foundational changes in society during the past century.

Of course new technologies rarely emerge in isolation. The industrial revolution is not just the story of harnessing steam power to factory production capability, but also the story of technological revolutions in transport, communication, construction, agriculture, resource extraction, and, of course, weapons development. These technological systems penetrated the innermost niches of society—home and family, school, workplace, community—and forced them to change. They also introduced completely new social phenomena, and stimulated the invention of completely new institutions.

The industrial revolution created the macroeconomic phenomenon of unemployment. Prior to the 19th century, even the most economically and politically advanced societies were dominantly agrarian and rural. For the majority of people, work was rooted in the home and the family. Vagaries of weather and transportation imposed irregularities and hardship, but most people and families harbored a diversity of skills that gave them independence from the marketplace and resilience to cope with a variety of challenges. In hard times, resort to subsistence farming and barter was usually possible.[3]

Industrialization and urbanization linked workers far more closely to the larger economic market, while removing the need and ability for them to maintain the diverse skills necessary for survival in the pre-industrial world. The traditional connection between manufacturing and agriculture in the home was sundered by new economic organization and by geography. Labor itself became a commodity, subject to the same fluctuations and influences as other commodities. During an economic downturn, factories fired people or closed down entirely. For the first time, workers could not easily respond to changing economic conditions by switching to a different type of work or moving to a subsistence mode. The political economist Karl Polanyi observed: "To separate labor from other activities of life and to subject it to the laws of the market was to annihilate all organic forms of existence and to replace them by a different type of organization, an atomistic and individualistic one."[4]

As technological innovation interacts with society to create new phenomena, such as unemployment, society also responds by developing new types of institutions and response mechanisms. Today we can recognize the problem of unemployment as central to a diversity of social, political, and economic structures and activities ranging from the organization of labor to insurance safety nets to educational programs to immigration policy. Unemployment rates are a key indicator of economic health, and a key determinant of political behavior. National and international economic policies focus strongly on managing unemployment, even as theoretical investigations seek to clarify the relation between unemployment rates and other key attributes of modern economies.

The general point is that transformational technology represents one variable in a complex assemblage of dynamic, interrelated societal activities. Decision making processes tend to address each of these activities in isolation from the others, e.g., conduct of research and development (R&D), dissemination of innovation products, development of regulations, reform of institutions. Concerted action occurs when a given innovation stimulates enough transformation to demand a response from other sectors of society. This response then triggers additional changes, which in turn demand further modulation. The process is reactive, discontinuous, disruptive, and sequential—like billiards. The challenge is to move toward a process of technology-supported societal progress where different sectors and activities can continually coevolve in response to knowledge about one another's needs and constraints—like an ecosystem. We are not there yet.

Transforming the Present

A brief consideration of evolution of information technologies helps to bring this look at societal transformation into the present. Gutenberg's perfection of the printing press of course had enormous transformational impact, allowing the broad dissemination of written texts and consequent expansion of information—and literacy—that undermined the Church's hegemony over knowledge and culture, and helped promote the dissolution of medieval social structure. Lewis Mumford suggested that the printed word represents "the media of reflective thought and deliberate action," a prerequisite, perhaps, for the intellectual achievements of the Enlightenment. But he also observed—as early as 1934—that new modes of electronic communication were increasing the speed of information exchange to levels that made reflection impossible, and increasing the volume of information transmission to a point that exceeded our absorptive capacity.[5]

The implications of the information and communication revolution on democracy itself are far from clear. On the one hand, proliferation of information dissemination networks means greater access by more people to more information—and a greater capacity to communicate one's ideas and preferences in democratic fora. Control of information by authoritarian governments is becoming increasingly futile, and organization of democratic opposition increasingly enhanced, by new information technologies. But when this same capacity translates into 10,000 identical e-mail messages sent to a Member of Congress in support of a particular bill, one is hard-pressed to suggest that democracy is the beneficiary. Of particular concern is the recent increase in public referenda aimed at bypassing the legislative process. The barriers to putting referenda on ballots have been enormously reduced by information and communication technologies that can be used to disseminate ideas and organize group action with relatively little effort. While on the one hand this type of direct democracy can be a refreshing antidote to sclerotic legislative process, on the other it is quite often devoid of any serious deliberative process or public discourse, reflecting perhaps the pique of one well-organized interest group or individual, and the substantiation of a Warholian politics where anyone with access to a decent list-serve can lead a movement for a day. Is democracy in transition?

The implications of the information and communications revolution on the distribution of economic benefits in society are also problematic. Does the troubling increase in wealth concentration that characterizes both the U.S. and the global economy derive from the way that advanced technologies diffuse in market economies? Does the synergistic character of information and communication networks mean that disenfranchised populations and nations will find it increasingly difficult to participate in the spectacular economic growth that we have seen in the past decade? In other words, are the benefits of technology becoming increasingly appropriable by particular sectors of society, and is this in part an attribute embodied in new types of technological systems? Society is ill-prepared to answer such questions, let alone act on them in a knowledgeable manner.

Paradoxically, concerns about appropriability cut both ways. In the information society, the increasing ease of information dissemination may also threaten our system for protecting intellectual property and innovation. From pirated CD's sold on the streets of Shanghai to the advent of Napster, the concept of intellectual property seems increasingly vulnerable. Are we looking to a future where such protection is no longer practically possible? Does a world without patents and copyrights seem unimaginable? More unimaginable than, say, the loss of monopoly over the written word would have seemed to the Church in 1450?

At issue here is not the value of change, but the path that change follows. What may look in retrospect like the march of progress may be experienced in real time as wrenching dislocation. The Dickensian squalor of 19th century London remains a symbol of the human impacts of technological change. Faced with unprecedented societal transformations, the English government (as well as other European states) failed to develop effective policies that could accommodate the rapid transition from rural agrarian to urban industrial society. Today, the plight of many overpopulated developing nations is the post-industrial, global manifestation of the same failure.

We see the fingerprints of societally-transforming technological systems in the controversy over genetically modified organisms; in the morally reprehensible situation where 24 million HIV-positive sub-Saharan Africans cannot possibly afford AIDS drugs that are widely available in the affluent world; in the 40 million Americans with no medical insurance; in the general inability of our public school systems to create a citizenry able to take advantage of the opportunities of the knowledge economy; in the challenges presented by the aging of our population; in the rising atmospheric carbon dioxide levels that reflect 150 years of industrial dynamism.

Even the unprecedented rise of civil and ethnic conflict throughout the world in the past decade can be plausibly connected to technological transformation. Approaching this phenomenon from entirely different directions, the political scientists Samuel Huntington and Benjamin Barber each conclude that advanced communication and information technologies have created new fora for expressing ethnic identity and pursuing and strengthening cultural solidarity. Virtual communities, for example, can act to maintain identity over great distance, while also more efficiently garnering resources to support the expression of cultural goals. As Barber observes: "Christian Fundamentalists [can] access Religion Forum on CompuServe Information Service while Muslims can surf the Internet until they find Mas'ood Cajee's Cybermuslim document." The result may be locally empowering and globally divisive.[6]

Nanotechnology and Societal Transformation

The marriage of science and technology beginning in the latter part of the 19th century accelerated the process of innovation, and thus the process of societal transformation as well. If the industrial revolution played itself out in less than 200 years, the electronics revolution seems likely to have a working life of perhaps 75 years, while the biotechnology revolution, although hardly yet on its feet, is already prophesied to be supplanted by (or perhaps to morph into) the nanotechnology revolution in the first half of the new century. What type of transformations might this revolution have in store?

Our point here is not to predict the future of nanotechnology and its impacts—an impossible task—but to illustrate the direction and scale of thinking that will be necessary if we are to successfully manage the interaction of new knowledge and innovation with society. Judging by the literature prepared by the government,[7] as well as the work of futurists and other techno-pundits,[8] the promise of nanotechnology to remake our world seems virtually infinite. So the first thing to say is that if—as is variously claimed—nanotechnology is going to revolutionize manufacturing, health care, travel, energy supply, food supply, and warfare, then it is going, as well, to transform labor and the workplace, the medical system, the transportation and power infrastructure, the agricultural enterprise, and the military. Each one of these technology-dependent sectors is operated by and for human beings, who act within institutions and cultures, according to particular regulations, norms, and heuristics, all of which may reflect decades or even centuries of evolution, negotiation, and tradition. Not one of them will be "revolutionized" without significant difficulty. The current chaos in our health care system is emblematic of this type of difficulty.

In the near term, the current state of knowledge may suggest that the first wave of useful nanotechnologies will lie in the area of detection and sensing. The capacity to detect, precisely identify, and perhaps isolate single molecules, viruses, or other complex, nanoscale structures has broad application in such areas as medical diagnosis, forensics, national defense, and environmental monitoring and control. The potential for direct benefits is obvious; how might this evolving capacity influence society?

When detection outpaces response capability—as it usually does—ethical and policy dilemmas inevitably arise. For example, it is already possible to identify genetic predisposition to certain diseases for which there are no known cures, or to diagnose congenital defects in fetuses for which the only cure is abortion. In the environmental realm, new technologies that detect pollutants at extremely low concentrations raise complex questions about risk thresholds and appropriate remediation standards. The presence of tiny amounts of toxic materials in groundwater may justifiably raise alarm among the public even if the health risk cannot be assessed, and the technological capacity for remediation does not exist. These types of dilemmas may be expected to accelerate and proliferate with the advance of nanodetection technologies.

Advances in sensing and detection may transform existing societal mechanisms and institutions that were designed to cope with uncertainty and incomplete or imprecise information. The insurance industry, for example, deals with incomplete knowledge about the health of specific individuals by spreading its risk among large populations. If there is no way to distinguish between someone who is going to suffer a potentially lethal middle-age heart attack, and someone who is going to live to 105, then they can both get health and life insurance. Society clearly gains from this arrangement: costs are broadly disseminated, and benefits are delivered to those who most need them.

Medical sensors that can, for example, "detect an array of medically relevant signals at high sensitivity and selectivity"[9] promise to aid diagnosis and treatment of disease, but also to develop predictive health profiles of individuals. Today, health and life insurance companies often use pre-existing conditions as a basis for denying or restricting coverage. The advent of nanodetection capabilities will considerably expand the information that insurance companies will want to use in making decisions about coverage. The generation of new information might thus destabilize the risk-spreading approach that allows equitable delivery of social benefits to broad populations. How will society respond?

Nanotechnology offers a dizzying range of potential benefits for military application. Recent history suggests that some of the earliest applications of nanotechnology will come in the military realm, where specific needs are well-articulated, and a customer—the Department of Defense—already exists. One area of desired nano-innovation lies in the "increased use of enhanced automation and robotics to offset reductions in military manpower, reduce risks to troops, and improve vehicle performance." (Budget, p. 20)[10] How might progress in this realm interact with the current trend toward rising civilian casualties (in absolute terms and relative to military personnel) in armed conflict worldwide? As increased robotic capability is realized in warfare, will we enter an era when it is safer to be a soldier in wartime than a civilian?

Such considerations are simple extrapolations of current trends in technological innovation and societal transformation. More adventurous speculation is tempting but is perhaps best confined to science fiction novels. The question of public response to nano-innovation, however, should not be avoided, even at this early stage. The ongoing experience of public opposition to old technologies such as nuclear power, new technologies such as genetically modified foods, and prospective technologies such as stem cell therapies, needs to be viewed as integral to the relationship between innovation and societal transformation.

Three observations are particularly relevant here. First, the impact of rapid technological innovation on people's lives is usually not consensual. Second, in the short term at least, the social changes induced by new technologies usually create both winners and losers (where what is lost may range from a job to an entire community). Third, rapid technological change can threaten the social structure, economic stability, and spiritual meaning that people strive in their lives to achieve. As the nanotechnology revolution begins to unfold in all its promise and diversity, such issues are bound to express themselves. They should not be viewed as threats, or as manifestations of intellectual weakness or repugnant ideology. Rather, they need to be recognized as a central part of the human context for technological change.

Preparing for the Revolution

> Now nanotechnology had made nearly anything possible, and so the cultural role in deciding what should be done with it had become far more important than imagining what could be done with it.
>
> —Neal Stephenson, *The Diamond Age*[11]

When resources are allocated for R&D programs, the implications for complex societal transformation are not considered. The fundamental assumption underlying the allocation process is that all societal outcomes will be positive, and that technological cause will lead directly to a desired societal effect. The literature promoting the National Nanotechnology Initiative expresses this view. The current policy approach thus addresses two elements:

- Conduct of Science and Technology

- Products of Science and Technology

These elements reflect the internal workings of the R&D enterprise. The fact that societal outcomes are not a serious part of the framework seems to derive from two beliefs: (1) that the science and technology enterprise has to be granted autonomy to chose its own direction of advance and innovation; and (2) that because we cannot predict the future of science or technological innovation, we cannot prepare for it in advance. These are oft-articulated arguments, not straw men. Yet the first is contradicted by reality, and the second is irrelevant. The direction of science and technology is in fact dictated by an enormous number of constraints (only one of which is the nature of nature itself). And preparation for the future obviously does not require accurate prediction; rather, it requires a foundation of knowledge upon which to base action, a capacity to learn from experience, close attention to what is going on in the present, and healthy and resilient institutions that can effectively respond or adapt to change in a timely manner.

If we flip the current S&T policy approach on its head, and start by thinking about desired social outcomes, rather than desired inputs to the R&D enterprise (i.e., more money), where would we begin? We might identify several very general categories of outcomes that most people would agree are worth thinking about. For example:

- Social equity: the distribution of the benefits of science and technology.

- Social purpose: the actual goals of societal development that we want to pursue or advance.

- Economic and Social enterprises: the shape and make-up of the institutions at the interface between technology and the human experience.

How can consideration of these types of outcomes be integrated into the S&T policy framework? The years since World War II have seen a very gradual evolution in the effort to connect thinking about S&T to thinking about the outcomes of S&T in society. A science policy report issued by the Truman Administration, for example, mentioned in its first pages the need to prepare for both the positive and negative impacts of scientific and technological change.[12] The rise of the environmental movement in the late 1960s reflected a public demand that society devote more S&T resources to the achievement of desired social outcomes like clean air and water. The creation of the congressional

Office of Technology Assessment reflected growing public concern about the need to understand the societal implications of technological choices. Over the past decade, federally funded programs on the human dimensions of global climate change, and the ethical, legal, and social implications of the human genome project and information technologies, have been supported as adjuncts to much, much larger core research agendas in the "hard" sciences. Yet S&T policy itself remains input-driven.

Concepts such as sustainability, and analytical tools such as human development indicators, provide conceptual frameworks for linking R&D to societal outcomes, and in fact imply that outcomes are to some degree implicit in the choices we make about R&D inputs. These types of insights point the way toward the next step: to implement an approach to R&D policy that addresses the complex interconnections between technological advance and societal response. Such an approach would need to integrate the pursuit of innovation with an evolving understanding of how innovation and society interact, and include mechanisms to feed this understanding back into the innovation process itself. (In a very specific way, the private sector does this as a matter of course, as it uses consumer input to continually refine and improve the next generation of products.)

If we wanted to be serious about preparing for the transformational power of a coming nanotechnology revolution, we would need first to get serious—at this very early stage—about developing knowledge and tools for more effectively connecting R&D inputs with desired societal outcomes. This in turn would require the creation of a dedicated intellectual, analytical, and institutional capability focused on understanding the dynamics of the science-society interface and feeding back into the evolving nanotechnology enterprise. Such a capability might include the following elements:

- *Analysis of past and current societal responses to transforming technologies.* A case history approach could be used to investigate the diverse avenues that society has followed in responding to a range of technological advances. Understanding the roles and relations between the media, academia, policy makers, institutions, and cultural factors could be the basis for assessing—and anticipating—the likely trajectories of technology-induced social change.

- *Comprehensive, real time assessment and monitoring of the nanoscience and nanotechnology enterprise.* At this relatively early stage, it should be feasible to build a database of important activities in nanotechnology, and then track the evolution of the enterprise over time, in terms of directions of research and innovation, resources used, public and private sector roles, publications and patents, marketed products, and other useful indicators. This type of information is essential to understanding potential impacts.

- *A science communication initiative, to foster dialogue among scientists, technologists, policy makers, the media, and the public.* Understanding, tracking, and enhancing the processes by which information about nanotechnology diffuses from the laboratory to the outside world is central to understanding the social transformation process as it occurs. Of equal importance is the need to understand and monitor how public attitudes and needs evolve, and how they reach back into the innovation system. Empirically grounded, research-based investigations on communication can be the basis for strategies to improve social choice in ways likely to secure favorable outcomes.

- *A constructive technology assessment process, with participants drawn from representatives of the R&D effort, the policy world, and the public.* Technology assessment is both a process for bringing together a range of relevant actors, and an evolving product that can inform and link the innovation and decision-making processes. Understanding the changing capabilities of both the nanotechnology enterprise and various sectors and institutions likely to be affected by the enterprise can contribute to a healthy policy making environment where innovation paths and social goals are compatible and mutually reinforcing.

Should nanoscience and nanotechnology yield even a small proportion of their anticipated advances, the impacts on society will be far-reaching and profound—"as socially transforming as the development of running water, electricity, antibiotics, and microelectronics."[13] We can allow these transformations to surprise and overwhelm us, and perhaps even threaten the prospects for further progress. Or we can choose to be smart about preparing for, understanding, responding to, and even managing the coming changes, in order to enhance the benefits, and reduce the disruption and dislocation, that must accompany any revolution.

Endnotes

1. White, Lynn T., *Medieval Technology and Social Change.* Oxford University Press, London, 1962.
2. Lemann, Nicholas. *The Promised Land: The Great Black Migration and How It Changed America.* Random House, New York, NY, 1991.
3. Keyssar, Alexander. *Out of Work: The First Century of Unemployment in Massachusetts.* Cambridge University Press, Cambridge, MA, 1986.
4. Polanyi, Karl. *The Great Transformation: The Political and Economic Origins of Our Time.* orig. ed. 1944, p. 163. Beacon Press, Boston, MA, 1957.
5. Mumford, Lewis. *Technics and Civilization.* p. 240. Harcourt Brace Jovanovich, Inc., New York, NY, 1934.
6. Barber, Benjamin. *Jihad vs. McWorld: How Globalism and Tribalism Are Reshaping the World.* p. 155-156. Ballantine Books, New York, NY 1996; Huntington, Samuel P. *The Clash of Civilizations and the Remaking of the World Order.* Simon and Schuster, New York, NY, 1996.
7. E.g., National Science and Technology Council. *Nanotechnology: Shaping the World Atom by Atom.* NSTC, Washington, DC, September 1999; National Science and Technology Council. *National Nanotechnology Initiative: Leading to the Next Industrial Revolution* (Supplement to the President's FY 2001 Budget) NSTC, Washington, DC, February 2000.
8. E.g., Cetron, M., and Davies, O., *Probable Tomorrows: How Science and Technology Will Transform Our Lives in the Next Twenty Years.* St. Martin's Press, New York, NY, 1997.
9. NSTC 2000, p. 45.
10. Ibid., p. 20.
11. Stephenson, Neal. *The Diamond Age.* Bantam Spectra, New York, NY, 2000.
12. Steelman, John R. *Science and Public Policy;* p. viii. U.S. Government Printing Office, Washington, DC, 1947.
13. NSTC 1999, p. 1.