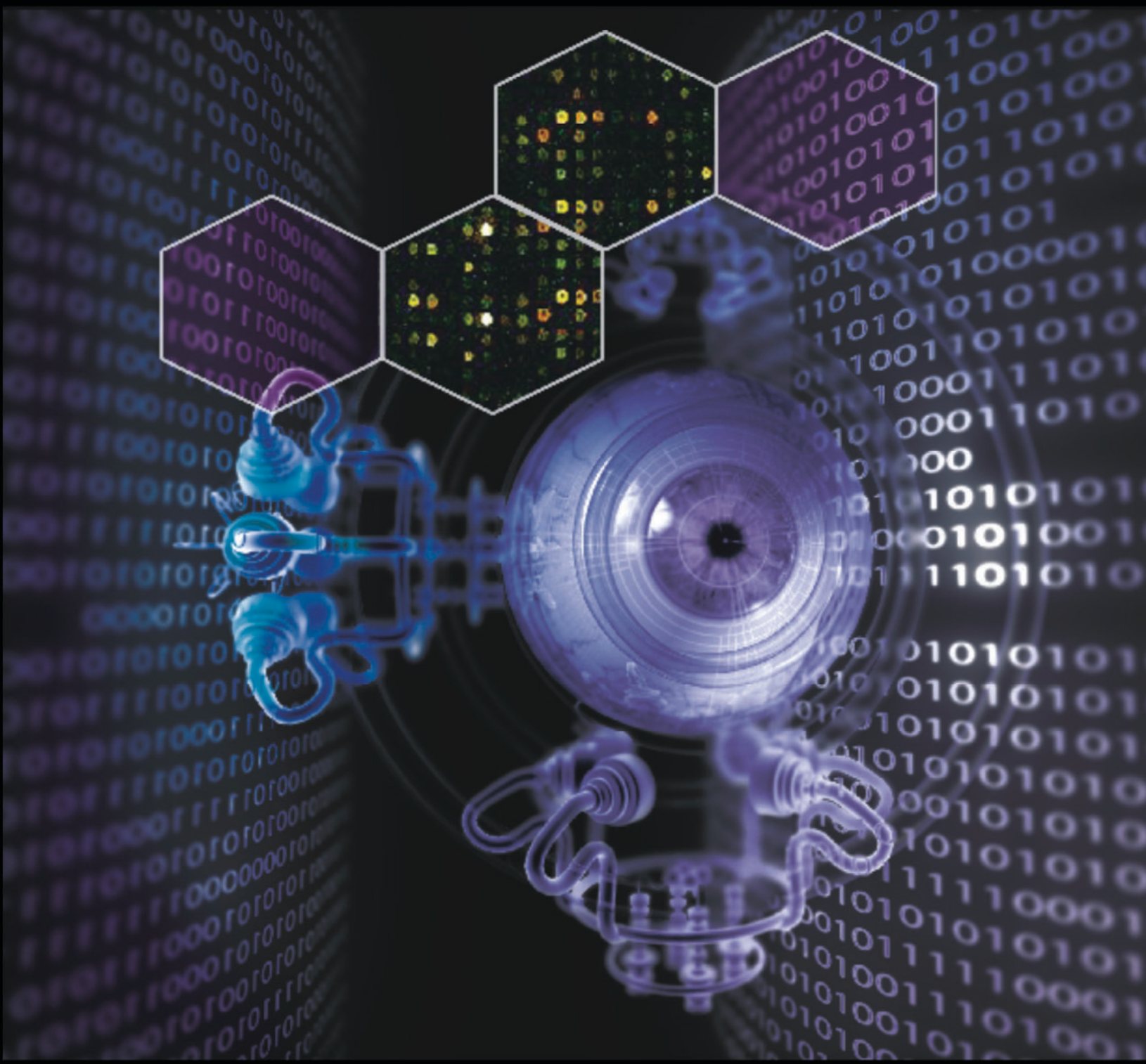
The Ray Kurzweil Reader

A collection of essays by Ray Kurzweil
published on KurzweilAI.net 2001–2003



Acknowledgements

The essays in this collection were published on KurzweilAI.net during 2001-2003, and have benefited from the devoted efforts of the KurzweilAI.net editorial team. Our team includes Amara D. Angelica, editor; Nanda Barker-Hook, editorial projects manager; Sarah Black, associate editor; Emily Brown, editorial assistant; and Celia Black-Brooks, graphics design manager and vice president of business development. Also providing technical and administrative support to KurzweilAI.net are Ken Linde, systems manager; Matt Bridges, lead software developer; Aaron Kleiner, chief operating and financial officer; Zoux, sound engineer and music consultant; Toshi Hoo, video engineering and videography consultant; Denise Scutellaro, accounting manager; Joan Walsh, accounting supervisor; Maria Ellis, accounting assistant; and Don Gonson, strategic advisor.

—Ray Kurzweil, Editor-in-Chief

TABLE OF CONTENTS

HOW TO BUILD A BRAIN 43

A machine is likely to achieve the ability of a human brain in the coming years. Ray Kurzweil has predicted that a $1,000 personal computer will match the computing speed and capacity of the human brain by around the year 2020. With human brain reverse engineering, we should have the software insights before 2030. This section explores the possibilities of machine intelligence and exotic new technologies for faster and more powerful computational machines, from cellular automata and DNA molecules to quantum computing. It also examines the controversial area of uploading your mind into a computer.

NANOTECHNOLOGY 197

Think small. The nanotechnology boom is beginning. Now how do we keep it under control?

Human Body Version 2.0

Ray Kurzweil

http://www.kurzweilai.net/articles/art0551.html

In the coming decades, a radical upgrading of our body's physical and mental systems, already underway, will use nanobots to augment and ultimately replace our organs. We already know how to prevent most degenerative disease through nutrition and supplementation; this will be a bridge to the emerging biotechnology revolution, which in turn will be a bridge to the nanotechnology revolution. By 2030, reverse-engineering of the human brain will have been completed and nonbiological intelligence will merge with our biological brains.

Published on KurzweilAI.net February 17, 2003. Ray Kurzweil presented a talk based on this article on February 21, 2003 at Time magazine's Future of Life Summit.

Sex has already been largely separated from its biological function. For the most part, we engage in sexual activity for intimate communication and sensual pleasure, not reproduction. Conversely, we have multiple methodologies for creating babies without physical sex, albeit most reproduction still does derive from the sex act. Although not condoned by all sectors of society, this disentanglement of sex from its biological function has been readily, even eagerly, adopted by the mainstream.

So why don't we provide the same extrication of purpose from biology for another activity that also provides both social intimacy and sensual pleasure, namely eating? We have crude ways of doing this today. Starch blockers, such as Bayer's Precose, partially prevent absorption of complex carbohydrates; fat blockers, such as Chitosan, bind to fat molecules, causing them to pass through the digestive tract; and sugar substitutes, such as Sucralose and Stevia, provide sweetness without calories. There are limitations and problems with each of these contemporary technologies, but a more effective generation of drugs is being developed that will block excess caloric absorption on the cellular level.

Let us consider, however, a more fundamental reengineering of the digestive process to disconnect the sensual aspects of eating from its original biological purpose: to provide nutrients into the bloodstream that are then delivered to each of our trillions of cells. These nutrients include caloric (energy-bearing) substances such as glucose (from carbohydrates), proteins, fats, and a myriad of trace molecules, such as vitamins, minerals, and phytochemicals, that provide building blocks and facilitating enzymes for diverse metabolic processes.

An Era of Abundance

Our knowledge of the complex pathways underlying digestive processes is rapidly expanding, although there is still a great deal we do not fully understand. On the one hand, digestion, like any other major human biological system, is astonishing in its intricacy and cleverness. Our bodies manage to extract the complex resources needed to survive, despite sharply varying conditions, while at the same time, filtering out a multiplicity of toxins.

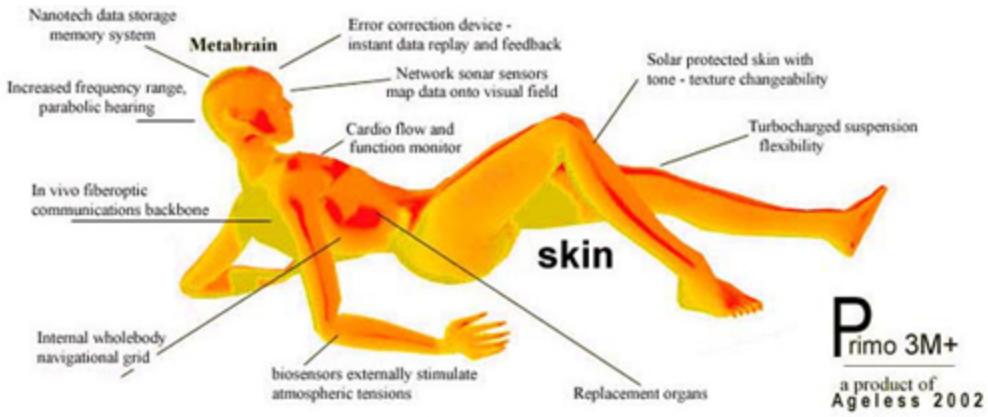
On the other hand, our bodies evolved in a very different era. Our digestive processes in particular are optimized for a situation that is dramatically dissimilar to the one we find ourselves in. For most of our biological heritage, there was a high likelihood that the next foraging or hunting season (and for a brief, relatively recent period, the next planting season) might be catastrophically lean. So it made sense for our bodies to hold on to every possible calorie. Today, this biological strategy is extremely counterproductive. Our outdated metabolic programming underlies our contemporary epidemic of obesity and fuels pathological processes of degenerative disease such as coronary artery disease, and type II diabetes.

Up until recently (on an evolutionary time scale), it was not in the interest of the species for old people like myself (I was born in 1948) to use up the limited resources of the clan. Evolution favored a short life span—life expectancy was 37 years only two centuries ago—so these restricted reserves could be devoted to the young, those caring for them, and those strong enough to perform intense physical work.

We now live in an era of great material abundance. Most work requires mental effort rather than physical exertion. A century ago, 30 percent of the U.S. work force worked on farms, with another 30 percent deployed in factories. Both of these figures are now under 3 percent. The significant majority of today's job categories, ranging from airline flight attendant to web designer, simply didn't exist a century ago. Circa 2003, we have the opportunity to continue to contribute to our civilization's exponentially growing knowledge base—incidentally, a unique attribute of our species—well past our child-rearing days.

Our species has already augmented the "natural" order of our life cycle through our technology: drugs, supplements, replacement parts for virtually all bodily systems, and many other interventions. We already have devices to replace our hips, knees, shoulders, elbows, wrists, jaws, teeth, skin, arteries, veins, heart valves, arms, legs, feet, fingers, and toes. Systems to replace more complex organs (for example, our hearts) are beginning to work. As we're learning the principles of operation of the human body and the brain, we will soon be in a position to design vastly superior systems that will be more enjoyable, last longer, and perform better, without susceptibility to breakdown, disease, and aging.

Artist and cultural catalyst Natasha Vita-More pioneered a conceptual design for one such system, called Primo Posthuman, designed for mobility, flexibility and superlongevity. It features innovations such as a metabrain for global-net connection with prosthetic neo-neocortex of AI interwoven with nanobots; smart skin that is solar protected with biosensors for tone and texture changeability, and high-acuity senses.



Introducing Human Body Version 2.0

We won't engineer human body version 2.0 all at once. It will be an incremental process, one already well under way. Although version 2.0 is a grand project, ultimately resulting in the radical upgrading of all our physical and mental systems, we will implement it one benign step at a time. Based on our current knowledge, we can already touch and feel the means for accomplishing each aspect of this vision.

From this perspective, let's return to a consideration of the digestive system. We already have a reasonably comprehensive picture of the constituent ingredients of the food we eat. We already have the means to survive without eating, using intravenous nutrition (for people who are unable to eat), although this is clearly not a pleasant process, given the current limitations in our technologies for getting substances in and out of the blood stream.

The next phase of improvement will be largely biochemical, in the form of drugs and supplements that will block excess caloric absorption and otherwise reprogram metabolic pathways for optimal health. We already have the knowledge to prevent most instances of degenerative disease, such as heart disease, stroke, type II diabetes, and cancer, through comprehensive programs of nutrition and supplementation, something which I personally do, and will describe in an upcoming book (*A Short Guide to a Long Life*, coauthored with Terry Grossman, M.D.). I view our current knowledge as a bridge to the full flowering of the biotechnology revolution, which in turn will be a bridge to the nanotechnology revolution.

It's All About Nanobots

In a famous scene from the movie, *The Graduate*, Benjamin's mentor gives him career advice in a single word: "plastics." Today, that word might be "software," or "biotechnology," but in another couple of decades, the word is likely to be "nanobots." Nanobots—blood-cell-sized robots—will provide the means to radically redesign our digestive systems, and, incidentally, just about everything else.

In an intermediate phase, nanobots in the digestive tract and bloodstream will intelligently extract the precise nutrients we need, call for needed additional nutrients and supplements

through our personal wireless local area network, and send the rest of the food we eat on its way to be passed through for elimination.

If this seems futuristic, keep in mind that intelligent machines are already making their way into our blood stream. There are dozens of projects underway to create blood-stream-based "biological microelectromechanical systems" (bioMEMS) with a wide range of diagnostic and therapeutic applications. BioMEMS devices are being designed to intelligently scout out pathogens and deliver medications in very precise ways.

For example, a researcher at the University of Illinois at Chicago has created a tiny capsule with pores measuring only seven nanometers. The pores let insulin out in a controlled manner but prevent antibodies from invading the pancreatic Islet cells inside the capsule. These nanoengineered devices have cured rats with type I diabetes, and there is no reason that the same methodology would fail to work in humans. Similar systems could precisely deliver dopamine to the brain for Parkinson's patients, provide blood-clotting factors for patients with hemophilia, and deliver cancer drugs directly to tumor sites. A new design provides up to 20 substance-containing reservoirs that can release their cargo at programmed times and locations in the body.

Kensall Wise, a professor of electrical engineering at the University of Michigan, has developed a tiny neural probe that can provide precise monitoring of the electrical activity of patients with neural diseases. Future designs are expected to also deliver drugs to precise locations in the brain. Kazushi Ishiyama at Tohoku University in Japan has developed micromachines that use microscopic-sized spinning screws to deliver drugs to small cancer tumors.

A particularly innovative micromachine developed by Sandia National Labs has actual microteeth with a jaw that opens and closes to trap individual cells and then implant them with substances such as DNA, proteins or drugs. There are already at least four major scientific conferences on bioMEMS and other approaches to developing micro- and nano-scale machines to go into the body and bloodstream.

Ultimately, the individualized nutrients needed for each person will be fully understood (including all the hundreds of phytochemicals) and easily and inexpensively available, so we won't need to bother with extracting nutrients from food at all. Just as we routinely engage in sex today for its relational and sensual gratification, we will gain the opportunity to disconnect the eating of food from the function of delivering nutrients into the bloodstream.

This technology should be reasonably mature by the 2020s. Nutrients will be introduced directly into the bloodstream by special metabolic nanobots. Sensors in our bloodstream and body, using wireless communication, will provide dynamic information on the nutrients needed at each point in time.

A key question in designing this technology will be the means by which these nanobots make their way in and out of the body. As I mentioned above, the technologies we have today, such as intravenous catheters, leave much to be desired. A significant benefit of nanobot technology is that unlike mere drugs and nutritional supplements, nanobots have a measure of intelligence. They can keep track of their own inventories, and intelligently slip in and out of our bodies in

clever ways. One scenario is that we would wear a special "nutrient garment" such as a belt or undershirt. This garment would be loaded with nutrient bearing nanobots, which would make their way in and out of our bodies through the skin or other body cavities.

At this stage of technological development, we will be able to eat whatever we want, whatever gives us pleasure and gastronomic fulfillment, and thereby unreservedly explore the culinary arts for their tastes, textures, and aromas. At the same time, we will provide an optimal flow of nutrients to our bloodstream, using a completely separate process. One possibility would be that all the food we eat would pass through a digestive tract that is now disconnected from any possible absorption into the bloodstream.

This would place a burden on our colon and bowel functions, so a more refined approach will dispense with the function of elimination. We will be able to accomplish this using special elimination nanobots that act like tiny garbage compactors. As the nutrient nanobots make their way from the nutrient garment into our bodies, the elimination nanobots will go the other way. Periodically, we would replace the nutrition garment for a fresh one. One might comment that we do obtain some pleasure from the elimination function, but I suspect that most people would be happy to do without it.

Ultimately we won't need to bother with special garments or explicit nutritional resources. Just as computation will eventually be ubiquitous and available everywhere, so too will basic metabolic nanobot resources be embedded everywhere in our environment. In addition, an important aspect of this system will be maintaining ample reserves of all needed resources inside the body. Our version 1.0 bodies do this to only a very limited extent, for example, storing a few minutes of oxygen in our blood, and a few days of caloric energy in glycogen and other reserves. Version 2.0 will provide substantially greater reserves, enabling us to be separated from metabolic resources for greatly extended periods of time.

Once perfected, we will no longer need version 1.0 of our digestive system at all. I pointed out above that our adoption of these technologies will be cautious and incremental, so we will not dispense with the old-fashioned digestive process when these technologies are first introduced. Most of us will wait for digestive system version 2.1 or even 2.2 before being willing to do dispense with version 1.0. After all, people didn't throw away their typewriters when the first generation of word processors was introduced. People held onto their vinyl record collections for many years after CDs came out (I still have mine). People are still holding onto their film cameras, although the tide is rapidly turning in favor of digital cameras.

However, these new technologies do ultimately dominate, and few people today still own a typewriter. The same phenomenon will happen with our reengineered bodies. Once we've worked out the inevitable complications that will arise with a radically reengineered gastrointestinal system, we will begin to rely on it more and more.

Programmable Blood

As we reverse-engineer (learn the principles of operation of) our various bodily systems, we will be in a position to engineer new systems that provide dramatic improvements. One pervasive system that has already been the subject of a comprehensive conceptual redesign is our blood.

One of the leading proponents of "nanomedicine," (redesigning our biological systems through engineering on a molecular scale) and author of a book with the same name is Robert Freitas, Research Scientist at nanotechnology firm Zyvex Corp. Freitas' ambitious manuscript is a comprehensive road map to rearchitecting our biological heritage. One of Freitas' designs is to replace (or augment) our red blood cells with artificial "respirocytes" that would enable us to hold our breath for four hours or do a top-speed sprint for 15 minutes without taking a breath. Like most of our biological systems, our red blood cells perform their oxygenating function very inefficiently, and Freitas has redesigned them for optimal performance. He has worked out many of the physical and chemical requirements in impressive detail.

It will be interesting to see how this development is dealt with in athletic contests. Presumably, the use of respirocytes and similar systems will be prohibited from Olympic contests, but then we will have the specter of teenagers in junior high school gymnasiums routinely outperforming Olympic athletes.

Freitas envisions micron-size artificial platelets that could achieve hemostasis (bleeding control) up to 1,000 times faster than biological platelets. Freitas describes nanorobotic microbivores (white blood cell replacements) that will download software to destroy specific infections hundreds of times faster than antibiotics, and that will be effective against all bacterial, viral and fungal infections, with no limitations of drug resistance.

I've personally watched (through a microscope) my own white blood cells surround and devour a pathogen, and I was struck with the remarkable sluggishness of this natural process. Although replacing our blood with billions of nanorobotic devices will require a lengthy process of development, refinement, and regulatory approval, we already have the conceptual knowledge to engineer substantial improvements over the remarkable but very inefficient methods used in our biological bodies.

Have a Heart, or Not

The next organ on my hit list is the heart. It's a remarkable machine, but it has a number of severe problems. It is subject to a myriad of failure modes, and represents a fundamental weakness in our potential longevity. The heart usually breaks down long before the rest of the body, and often very prematurely.

Although artificial hearts are beginning to work, a more effective approach will be to get rid of the heart altogether. Among Freitas' designs are nanorobotic blood cell replacements that provide their own mobility. If the blood system moves with its own movement, the engineering issues of the extreme pressures required for centralized pumping can be eliminated. As we

perfect the means of transferring nanobots to and from the blood supply, we can also continuously replace the nanobots comprising our blood supply.

Energy will be provided by microscopic-sized hydrogen fuel cells. Integrated Fuel Cell Technologies, one of many companies pioneering fuel cell technology, has already created microscopic-sized fuel cells. Their first-generation design provides tens of thousands of fuel cells on an integrated circuit and is intended to power portable electronics.

With the respirocytes providing greatly extended access to oxygenation, we will be in a position to eliminate the lungs by using nanobots to provide oxygen and remove carbon dioxide. One might point out that we take pleasure in breathing (even more so than elimination!). As with all of these redesigns, we will certainly go through intermediate stages where these technologies augment our natural systems, so we can have the best of both worlds. Eventually, however, there will be no reason to continue with the complications of actual breathing and the requirement of having breathable air everywhere we go. If we really find breathing that pleasurable, we will develop virtual ways of having this sensual experience.

We also won't need the various organs that produce chemicals, hormones, and enzymes that flow into the blood and other metabolic pathways. We already create bio-identical versions of many of these substances, and we will have the means to routinely create all biochemically relevant substances within a couple of decades. These substances (to the extent that we still need them) will be delivered via nanobots, controlled by intelligent biofeedback systems to maintain and balance required levels, just as our "natural" systems do today (for example, the control of insulin levels by the pancreatic Islet cells). Since we are eliminating most of our biological organs, many of these substances may no longer be needed, and will be replaced by other resources that are required by the nanorobotic systems.

Similarly the organs that filter the blood for impurities, such as the kidneys, can also be replaced by nanorobot-based elimination services.

It is important to emphasize that this redesign process will not be accomplished in a single design cycle. Each organ and each idea will have its own progression, intermediate designs, and stages of implementation. Nonetheless, we are clearly headed towards a fundamental and radical redesign of the extremely inefficient and limited functionality of human body version 1.0.

So What's Left?

Let's consider where we are. We've eliminated the heart, lungs, red and white blood cells, platelets, pancreas, thyroid and all the hormone-producing organs, kidneys, bladder, liver, lower esophagus, stomach, small intestines, large intestines, and bowel. What we have left at this point is the skeleton, skin, sex organs, mouth and upper esophagus, and brain.

The skeleton is a stable structure, and we already have a reasonable understanding of how it works. We replace parts of it today, although our current technology for doing this has severe limitations. Interlinking nanobots will provide the ability to augment and ultimately replace the skeleton. Replacing portions of the skeleton today requires painful surgery, but replacing it

through nanobots from within can be a gradual and noninvasive process. The human skeleton version 2.0 will very strong, stable, and self repairing.

We will not notice the absence of many of our organs, such as the liver and pancreas, as we do not directly experience their functionality. The skin, however, is an organ we will actually want to keep, or at least we will want to maintain its functionality. The skin, which includes our primary and secondary sex organs, provides a vital function of communication and pleasure. Nonetheless, we will ultimately be able to improve on the skin with new nanoengineered supple materials that will provide greater protection from physical and thermal environmental effects while enhancing our capacity for intimate communication and pleasure. The same observation holds for the mouth and upper esophagus, which comprise the remaining aspects of the digestive system that we use to experience the act of eating.

Redesigning the Human Brain

The process of reverse engineering and redesign will also encompass the most important system in our bodies: the brain. The brain is at least as complex as all the other organs put together, with approximately half of our genetic code devoted to its design. It is a misconception to regard the brain as a single organ. It is actually an intricate collection of information-processing organs, interconnected in an elaborate hierarchy, as is the accident of our evolutionary history.

The process of understanding the principles of operation of the human brain is already well under way. The underlying technologies of brain scanning and neuron modeling are scaling up exponentially, as is our overall knowledge of human brain function. We already have detailed mathematical models of a couple dozen of the several hundred regions that comprise the human brain.

The age of neural implants is also well under way. We have brain implants based on "neuromorphic" modeling (i.e., reverse-engineering of the human brain and nervous system) for a rapidly growing list of brain regions. A friend of mine who became deaf while an adult can now engage in telephone conversations again because of his cochlear implant, a device that interfaces directly with the auditory nervous system. He plans to replace it with a new model with a thousand levels of frequency discrimination, which will enable him to hear music once again. He laments that he has had the same melodies playing in his head for the past 15 years and is looking forward to hearing some new tunes. A future generation of cochlear implants now on the drawing board will provide levels of frequency discrimination that go significantly beyond that of "normal" hearing.

Researchers at MIT and Harvard are developing neural implants to replace damaged retinas. There are brain implants for Parkinson's patients that communicate directly with the ventral posterior nucleus and subthalmic nucleus regions of the brain to reverse the most devastating symptoms of this disease. An implant for people with cerebral palsy and multiple sclerosis communicates with the ventral lateral thalamus and has been effective in controlling tremors. "Rather than treat the brain like soup, adding chemicals that enhance or suppress certain neurotransmitters," says Rick Trosch, an American physician helping to pioneer these therapies, "we're now treating it like circuitry."

A variety of techniques are being developed to provide the communications bridge between the wet analog world of biological information processing and digital electronics. Researchers at Germany's Max Planck Institute have developed noninvasive devices that can communicate with neurons in both directions. They demonstrated their "neuron transistor" by controlling the movements of a living leech from a personal computer. Similar technology has been used to reconnect leech neurons and to coax them to perform simple logical and arithmetic problems. Scientists are now experimenting with a new design called "quantum dots," which uses tiny crystals of semiconductor material to connect electronic devices with neurons.

These developments provide the promise of reconnecting broken neural pathways for people with nerve damage and spinal cord injuries. It has long been thought that recreating these pathways would only be feasible for recently injured patients because nerves gradually deteriorate when unused. A recent discovery, however, shows the feasibility of a neuroprosthetic system for patients with long-standing spinal cord injuries. Researchers at the University of Utah asked a group of long-term quadriplegic patients to move their limbs in a variety of ways and then observed the response of their brains, using magnetic resonance imaging (MRI). Although the neural pathways to their limbs had been inactive for many years, the pattern of their brain activity when attempting to move their limbs was very close to that observed in non-disabled persons.

We will, therefore, be able to place sensors in the brain of a paralyzed person (e.g., Christopher Reeve) that will be programmed to recognize the brain patterns associated with intended movements and then stimulate the appropriate sequence of muscle movements. For those patients whose muscles no longer function, there are already designs for "nanoelectromechanical" systems (NEMS) that can expand and contract to replace damaged muscles and that can be activated by either real or artificial nerves.

We Are Becoming Cyborgs

We are rapidly growing more intimate with our technology. Computers started out as large remote machines in air-conditioned rooms tended by white-coated technicians. Subsequently they moved onto our desks, then under our arms, and now in our pockets. Soon, we'll routinely put them inside our bodies and brains. Ultimately we will become more nonbiological than biological.

The compelling benefits in overcoming profound diseases and disabilities will keep these technologies on a rapid course, but medical applications represent only the early adoption phase. As the technologies become established, there will be no barriers to using them for the expansion of human potential. In my view, expanding our potential is precisely the primary distinction of our species.

Moreover, all of the underlying technologies are accelerating. The power of computation has grown at a double exponential rate for all of the past century, and will continue to do so well into this century through the power of three-dimensional computing. Communication bandwidths and the pace of brain reverse-engineering are also quickening. Meanwhile, according to my models,

the size of technology is shrinking at a rate of 5.6 per linear dimension per decade, which will make nanotechnology ubiquitous during the 2020s.

By the end of this decade, computing will disappear as a separate technology that we need to carry with us. We'll routinely have high-resolution images encompassing the entire visual field written directly to our retinas from our eyeglasses and contact lenses (the Department of Defense is already using technology along these lines from Microvision, a company based in Bothell, Washington). We'll have very-high-speed wireless connection to the Internet at all times. The electronics for all of this will be embedded in our clothing. Circa 2010, these very personal computers will enable us to meet with each other in full-immersion, visual-auditory, virtual-reality environments as well as augment our vision with location- and time-specific information at all times.

By 2030, electronics will utilize molecule-sized circuits, the reverse-engineering of the human brain will have been completed, and bioMEMS will have evolved into bioNEMS (biological *nano*electromechanical systems). It will be routine to have billions of nanobots (nano-scale robots) coursing through the capillaries of our brains, communicating with each other (over a wireless local area network), as well as with our biological neurons and with the Internet. One application will be to provide full-immersion virtual reality that encompasses all of our senses. When we want to enter a virtual-reality environment, the nanobots will replace the signals from our real senses with the signals that our brain would receive if we were actually in the virtual environment.

We will have a panoply of virtual environments to choose from, including earthly worlds that we are familiar with, as well as those with no earthly counterpart. We will be able to go to these virtual places and have any kind of interaction with other real (as well as simulated) people, ranging from business negotiations to sensual encounters. In virtual reality, we won't be restricted to a single personality, since we will be able to change our appearance and become other people.

Experience Beamers

"Experience beamers" will beam their entire flow of sensory experiences as well as the neurological correlates of their emotional reactions out on the Web just as people today beam their bedroom images from their web cams. A popular pastime will be to plug in to someone else's sensory-emotional beam and experience what it's like to be someone else, à la the plot concept of the movie "Being John Malkovich." There will also be a vast selection of archived experiences to choose from. The design of virtual environments and the creation of archived full-immersion experiences will become new art forms.

The most important application of circa-2030 nanobots will be to literally expand our minds. We're limited today to a mere hundred trillion interneuronal connections; we will be able to augment these by adding virtual connections via nanobot communication. This will provide us with the opportunity to vastly expand our pattern recognition abilities, memories, and overall thinking capacity as well as directly interface with powerful forms of nonbiological intelligence.

It's important to note that once nonbiological intelligence gets a foothold in our brains (a threshold we've already passed), it will grow exponentially, as is the accelerating nature of information-based technologies. A one-inch cube of nanotube circuitry (which is already working at smaller scales in laboratories) will be at least a million times more powerful than the human brain. By 2040, the nonbiological portion of our intelligence will be far more powerful than the biological portion. It will, however, still be part of the human-machine civilization, having been derived from human intelligence, i.e., created by humans (or machines created by humans) and based at least in part on the reverse-engineering of the human nervous system.

Stephen Hawking recently commented in the German magazine *Focus* that computer intelligence will surpass that of humans within a few decades. He advocated that we "develop as quickly as possible technologies that make possible a direct connection between brain and computer, so that artificial brains contribute to human intelligence rather than opposing it." Hawking can take comfort that the development program he is recommending is well under way.

Human Cloning is the Least Interesting Application of Cloning Technology

Ray Kurzweil

http://www.kurzweilai.net/articles/art0535.html

Cloning is an extremely important technology—not for cloning humans but for life extension: therapeutic cloning of one's own organs, creating new tissues to replace defective tissues or organs, or replacing one's organs and tissues with their "young" telomere-extended replacements without surgery. Cloning even offers a possible solution for world hunger: creating meat without animals.

Published on KurzweilAI.net January 4, 2003.

All responsible ethicists, including this author, consider human cloning at the present time to be unethical. The reasons have nothing to do with the slippery (slope) issues of manipulating human life. Rather, the technology today simply does not work reliably. The current technique of fusing a cell nucleus from a donor to an egg cell using an electric spark causes a high level of genetic errors.

This is the primary reason that most of the fetuses created in this way do not make it to term. Those that do nonetheless have genetic defects. Dolly developed an obesity problem in adulthood, and the majority of the cloned animals produced thus far have had unpredictable health problems.

Scientists have a number of ideas for perfecting this process, including alternative ways of fusing the nucleus and egg cell, but until the technology is demonstrably safe, it would be unethical to create a human life with such a high likelihood of severe health problems.

Regardless of whether or not the recent announcement of a cloned baby turns out to be legitimate, there is no doubt that human cloning will occur, and occur soon, driven by all the usual reasons, ranging from its publicity value to its utility as a very weak form of immortality. The methods that are demonstrable in advanced animals will work quite well in humans. Once the technology is perfected in terms of safety, the ethical barriers will be feeble if they exist at all.

In my view, cloning is an extremely important technology, but the cloning of humans is the least of it. Let me first address its most valuable applications and then return to its most controversial one.

The early 21st century will be shaped by accelerating and interacting technological transformations, all based in one way or another on information. These include the explicit information technologies of intelligent machines, robotics, nanotechnology, and virtual reality.

Of perhaps even more immediate impact on human longevity and well-being will be the multiple and intersecting biological revolutions, which are based on understanding the information processes underlying life and disease, such as rational drug design, genomics, proteomics, and genetic cloning.

Why is cloning important?

The most immediate application of cloning is improved breeding by being able to directly reproduce an animal with a desirable set of genetic traits. A powerful example is reproducing animals from transgenic embryos (embryos with foreign genes) for pharmaceutical production. A case in point: one of the most promising new anti-cancer treatments is an antiangiogenesis drug (a drug that inhibits tumors from creating the new capillary networks needed for their growth) called aaATIII, which is produced in the milk of transgenic goats.

Another exciting application is recreating animals from endangered species. By cryopreserving cells from these species, they never need become extinct. It will eventually be possible to recreate animals from recently extinct species. This past year, scientists were able to synthesize DNA for the Tasmanian Tiger, which has been extinct for 65 years, with the hope of bringing this species back to life. As for long extinct species (e.g., dinosaurs), there is a high level of doubt that we will find the fully intact DNA required in a single preserved cell, but it is quite possible that we will eventually be able to synthesize the DNA needed by patching together the information derived from multiple inactive fragments.

Therapeutic cloning

Another valuable emerging application is therapeutic cloning of one's own organs. Here we don't clone the entire person (you), but rather directly create one of your organs. By starting with germ line cells, differentiation (into different types of cells) is triggered prior to the formation of a fetus. Because differentiation takes place during the pre-fetal stage (i.e., prior to implantation of a fetus), most ethicists believe that this process does not raise ethical concerns, although this issue has been highly contentious.
Another highly promising approach is "human somatic cell engineering," which bypasses fetal stem cells entirely. These emerging technologies create new tissues with a patient's own DNA by modifying one type of cell (such as a skin cell) directly into another (such as a pancreatic Islet cell or a heart cell) without the use of fetal stem cells. There have been breakthroughs in this area in the past year. For example, scientists from the U.S. and Norway successfully converted human skill cells directly into immune system cells and nerve cells.

Consider the question: What is the difference between a skin cell and any other type of cell in the body? After all, they all have the same DNA. The differences are found in protein signaling factors that we are now beginning to understand. By manipulating these proteins, we can trick one type of cell into becoming another.

Perfecting this technology would not only diffuse a contentious ethical and political issue, it is also the ideal solution from a scientific perspective. If I need pancreatic Islet cells, or kidney

tissues—or even whole new heart—to avoid autoimmune reactions, I would strongly prefer to obtain these with my own DNA, not the DNA from someone else's germ line cells.

This process will directly grow an organ with your genetic makeup. Perhaps most importantly, the new organ has its telemeres (the chemical "beads" at the end of DNA that get shorter every time a cell divides) fully extended to their original youthful length, so that the new organ is effectively young again. So an 80-year-old man could have his heart replaced with his own "25-year-old" heart.

The injection of pancreatic Islet cells is already showing great promise in treating type I Diabetes, but contemporary treatments require strong anti-rejection drugs, and the availability of these cells for transplantation is very limited. With this type of somatic cell engineering, a type I Diabetic will be able to produce his own Islet cells with his own genetic makeup, eliminating both the rejection and availability problems and thereby curing his Diabetes.

Even more exciting is the prospect of replacing one's organs and tissues with their "young" telomere-extended replacements without surgery. By introducing cloned telomere-extended cells into an organ, these cells will integrate themselves with the older cells. By repeated treatments of this kind over a period of time, the organ will end up being dominated by the younger cells. We normally replace our own cells on a regular basis anyway, so why not do so with youthful telomere-extended cells rather than telomere-shortened ones? There's no reason why we couldn't do this with every organ and tissue in our body. We would thereby grow progressively younger.

Solving world hunger

Another exciting opportunity is to create meat without animals. As with therapeutic cloning, we would not be creating the entire animal, but rather directly producing the desired animal parts or flesh. Essentially, all of the meat—billions of pounds of it—would in essence be from a single animal. What's the point of doing this? For one thing, we could eliminate human hunger.

By creating meat in this way, it becomes subject to the "law of accelerating returns," which is the exponential improvements in price-performance of information based technologies over time. So meat produced in this way will ultimately be extremely inexpensive. It could cost less than one percent of conventionally produced meat. Even though hunger in the world today is certainly exacerbated by political issues and conflicts, meat will become so inexpensive that it will have a profound effect on the affordability of food.

The advent of animal-less meat will also eliminate animal suffering. The economics of factory farming place a very low priority on the comfort and life style of the animals. They are essentially cogs in a machine, and suffer on a massive scale. Although animal activists may prefer that everyone become a vegetarian, that is not likely, and some research suggests would not be ideal for everyone from a nutritional perspective. With animal-less meat, there would be no animal suffering. We could use the same approach for such animal byproducts as leather, and, dare I say, fur. The enormous ecological damage created by factory farming would also be eliminated. And we could produce meat with a far more desirable nutritional profile.

Which brings us again to human cloning, in my mind the least interesting application. Once the technology is perfected (which is not the case today), I see neither the acute ethical dilemmas nor the profound promise that ethicists and enthusiasts have debated. So we'll have genetic twins separated by one or more generations: it's the sort of idea society absorbs in its sleep. It's far different from mental cloning in which a person's entire personality, memory, skills, and history will ultimately be downloaded into a different, and most likely more powerful, thinking medium. There's no issue of philosophical identity with genetic cloning-genetic clones are different people, even more so than conventional twins today.

But if we consider the full concept of cloning from cell to organisms, the benefits have enormous synergy with the other revolutions occurring in biology as well as in computer technology. As we learn to understand the genome of both humans and animals, and as we develop powerful new means of harnessing genetic information, cloning provides the means to replicate animals, organs, and cells. And that has profound implications for health and well-being, of both ourselves and our evolutionary cousins in the animal kingdom.

Dialogue Between Ray Kurzweil, Eric Drexler, and Robert Bradbury

Ray Kurzweil, Eric Drexler, Robert Bradbury

http://www.kurzweilai.net/articles/art0533.html

What would it take to achieve successful cryonics reanimation of a fully functioning human brain, with memories intact? A conversation at the recent Alcor Conference on Extreme Life Extension between Ray Kurzweil and Eric Drexler sparked an email discussion of this question. They agreed that despite the challenges, the brain's functions and memories can be represented surprisingly compactly, suggesting that successful reanimation of the brain may be achievable.

Published on KurzweilAI.net December 3, 2002. E-mail dialogue on November 23, 2002. Comments by Robert Bradbury added January 15, 2003.

Ray Kurzweil: Eric, I greatly enjoyed our brief opportunity to share ideas (difficulty of adding bits to quantum computing, cryonics reanimation, etc.). Also, it was exciting to hear your insightful perspective on the field you founded, now that it's gone—from what was regarded in the mainstream anyway as beyond-the-fringe speculation—to, well, mainstream science and engineering.

I had a few questions and/or comments (depending on whether I'm understanding what you said correctly). Your lecture had a very high idea density, so I may have misheard some details.

With regard to cryonics reanimation, I fully agree with you that preserving structure (i.e., information) is the key requirement, that it is not necessary to preserve cellular functionality. I have every confidence that nanobots will be able to go in and fix every cell, indeed every little machine in every cell. The key is to preserve the information. And I'll also grant that we could lose some of the information; after all, we lose some information every day of our lives anyway. But the primary information needs to be preserved. So we need to ask, what are the types of information required?

One is to identify the neuron cells, including their type. This is the easiest requirement. Unless the cryonics process has made a complete mess of things, the cells should be identifiable. By the time reanimation is feasible, we will fully understand the types of neurons and be able to readily identify them from the slightest clues. These neurons (or their equivalents) could then all be reconstructed.

The second requirement is the interconnections. This morphology is one key aspect of our knowledge and experience. We know that the brain is continually adding and pruning connections; it's a primary aspect of its learning and self-organizing principle of operation. The interconnections are much finer than the neurons themselves (for example, with current brain imaging techniques, we can typically see the neurons but we do not yet clearly see the

interneuronal connections). Again, I believe it's likely that this can be preserved, provided that the vitrification has been done quickly enough. It would not be necessary that the connections be functional or even fully evident, as long as it can be inferred where they were. And it would be okay if some fraction were not identifiable.

It's the third requirement that concerns me; the neurotransmitter concentrations, which are contained in structures that are finer yet than the interneuronal connections. These are, in my view, also critical aspects of the brain's learning process. We see the analogue of the neurotransmitter concentrations in the simplified neural net models that I use routinely in my pattern recognition work. The learning of the net is reflected in the connection weights as well as the connection topology (some neural net methods allow for self-organization of the topology, some do not, but all provide for self-organization of the weights). Without the weights, the net has no competence.

If the very-fine-resolution neurotransmitter concentrations are not identifiable, the downside is not equivalent to merely an amnesia patient who has lost his memory of his name, profession, family members, etc. Our learning, reflected as it is in both interneuronal connection topology and neurotransmitter concentration patterns, underlies knowledge that is far broader than these routine forms of memory, including our "knowledge" of language, how to think, how to recognize objects, how to eat, how to walk and perform all of our skills, etc. Loss of this information would result in a brain with no competence at all. It would be worse than a newborn's brain, which is at least designed to begin reorganizing itself. A brain with the connections intact but none of the neurotransmitter concentrations would have no competence of any kind and a connection pattern that would be too specific to relearn all of these skills and basic knowledge.

It's not clear whether the current vitrification-preservation process maintains this vital type of information. We could readily conduct an experiment to find out. We could vitrify the brain of a mouse and then do a destructive scan while still vitrified to see if the neurotransmitter concentrations are still evident. We could also confirm that the connections are evident as well.

The type of long-term memory that an amnesia patient has lost is just one type of knowledge in the brain. At the deepest level, the brain's self-organizing paradigm underlies our knowledge and all competency that we have gained since our fetal days (even prior to birth).

As a second issue, you said something about it being sufficient to just have preserved the big toe or the nose to reconstruct the brain. I'm not sure what you meant by that. Clearly none of the brain structure is revealed by body parts outside the brain. The only conceivable way one could restore a brain from the toe would be from the genome, which one can discover from any cell. And indeed, one could grow a brain from the genome. This would be, however, a fetal brain, which is a genetic clone of the original person, equivalent to an identical twin (displaced in time). One could even provide a learning and maturing experience for this brain in which the usual 20 odd years were sped up to 20 days or less, but this would still be just a biological clone, not the original person.

Finally, you said (if I heard you correctly) that the amount of information in the brain (presumably needed for reanimation) is about 1 gigabyte. My own estimates are quite different. It is true that genetic information is very low, although as I discussed above, genetic information is not at all sufficient to recreate a person. The genome has about 0.8 gigabytes of information. There is massive redundancy, however. For example, the sequence "ALU" is repeated 300,000 times. If one compresses the genome using standard data compression to remove redundancy, estimates are that one can achieve about 30 to 1 lossless compression, which brings us down to about 25 megabytes. About half of that comprises the brain, or about 12 megabytes. That's the initial design plan.

If we consider the amount of information in a mature human brain, however, we have about 10^{11} neurons with 10^3 average fan-out of connections, for an estimated total of 10^{14} connections. For each connection, we need to specify (i) the neurons that this connection is connected to, (ii) some information about its pathway as the pathway affects analog aspects of its electrochemical information processing, and (iii) the neurotransmitter concentrations in associated synapses. If we estimate about 10^2 bytes of information to encode these details (which may be low), we have 10^{16} bytes, considerably more than the 10^9bytes that you mentioned.

One might ask: How do we get from 10^7 bytes that specify the brain in the genome to 10^{16} bytes in the mature brain? This is not hard to understand, since we do this type of meaningful data expansion routinely in our self-organizing software paradigms. For example, a genetic algorithm can be efficiently coded, but in turn creates data far greater in size than itself using a stochastic process, which in turn self-organizes in response to a complex environment (the problem space). The result of this process is meaningful information far greater than the original program. We know that this is exactly how the creation of the brain works. The genome specifies initially semi-random interneuronal connection wiring patterns in specific regions of the brain (random within certain constraints and rules), and these patterns (along with the neurotransmitter-concentration levels) then undergo their own internal evolutionary process to self-organize to reflect the interactions of that person with their experiences and environment. That is how we get from 10^7 bytes of brain specification in the genome to 10^{16} bytes of information in a mature brain. I think 10^9 bytes is a significant underestimate of the amount of information required to reanimate a mature human brain.

I'd be interested in your own reflections on these thoughts, with my best wishes.

Eric Drexler: Ray—Thanks for your comments and questions. Our thinking seems closely parallel on most points.

Regarding neurotransmitters, I think it is best to focus not on the molecules themselves and their concentrations, but rather on the machinery that synthesizes, transports, releases, senses, and recycles them. The state of this machinery must closely track long-term functional changes (i.e, long-term memory or LTM), and much of this machinery is an integral part of synaptic structure.

Regarding my toe-based reconstruction scenario [creating a brain from a bit of tissue containing intact DNA—Ed.], this is indeed no better than genetically based reconstruction together with loading of more-or-less default skills and memories—corresponding to a peculiar but profound

state of amnesia. My point was merely that even this worst-case outcome is still what modern medicine would label a success: the patient walks out the door in good health. (Note that neurosurgeons seldom ask whether the patient who walks out is "the same patient" as the one who walked in.) Most of us wouldn't look forward to such an outcome, of course, and we expect much better when suspension occurs under good conditions.

Information-theoretic content of long-term memory

Regarding the information content of the brain, both the input and output data sets for reconstruction must indeed be vastly larger than a gigabyte, for the reasons you outline. The lower number [10^9] corresponds to an estimate of the information-theoretic content of human long term memory found (according to Marvin Minsky) by researchers at Bell Labs. They tried various methods to get information into and out of human LTM, and couldn't find learning rates above a few bits per second. Integrated over a lifespan, this
yields the above number. If this is so, it suggests that information storage in the brain is indeed massively redundant, perhaps for powerful function-enabling reasons. (Identifying redundancy this way, of course, gives no hint of how to construct a compression and decompression algorithm.)

Best wishes, with thanks for all you've done.

P.S. A Google search yields a discussion of the Bell Labs result by, yes, Ralph Merkle.

Ray Kurzweil: Okay, I think we're converging on some commonality.

On the neurotransmitter concentration level issue, you wrote: "Regarding neurotransmitters, I think it is best to focus not on the molecules themselves and their concentrations, but rather on the machinery that synthesizes, transports, releases, senses, and recycles them. The state of this machinery must closely track long-term functional changes (i.e, LTM), and much of this machinery is an integral part of synaptic structure."

I would compare the "machinery" to any other memory machinery. If we have the design for a bit of memory in a DRAM system, then we basically know the mechanics for the other bits. It is true that in the brain there are hundreds of different mechanisms that we could call memory, but each of these mechanisms is repeated many millions of times. This machinery, however, is not something we would need to infer from the preserved brain of a suspended patient. By the time reanimation is feasible, we will have long since reverse-engineered these basic mechanisms of the human brain, and thus would know them all. What we do need specifically for a particular patient is the state of that person's memory (again, memory referring to all skills). The state of my memory is not the same as that of someone else, so that is the whole point of preserving my brain.

And that state is contained in at least two forms: the interneuronal connection patterns (which we know is part of how the brain retains knowledge and is not a fixed structure) and the neurotransmitter concentration levels in the approximately 10^{14} synapses.

My concern is that this memory state information (particularly the neurotransmitter concentration levels) may not be retained by current methods. However, this is testable right now. We don't have to wait 40 to 50 years to find this out. I think it should be a high priority to do this experiment on a mouse brain as I suggested above (for animal lovers, we could use a sick mouse).

You appear to be alluding to a somewhat different approach, which is to extract the "LTM," which is likely to be a far more compact structure than the thousands of trillions of bytes represented by the connection and neurotransmitter patterns (CNP). As I discuss below, I agree that the LTM is far more compact. However, we are not extracting an efficient LTM during cryopreservation, so the only way to obtain it during cryo reanimation would be to retain its inefficient representation in the CNP.

You bring up some interesting and important issues when you wrote, "Regarding my toe-based reconstruction scenario, this is indeed no better than genetically-based reconstruction together with loading of more-or-less default skills and memories—corresponding to a peculiar but profound state of amnesia. My point was merely that even this worst-case outcome is still what modern medicine would label a success: the patient walks out the door in good health."

I agree that this would be feasible by the time reanimation is feasible. The means for "loading" these "default skills and memories" is likely to be along the lines that I described above, to use "a learning and maturing experience for this brain in which the usual 20 odd years were sped up to 20 days or less." Since the human brain as currently designed does not allow for explicit "loading" of memories and skills, these attributes need to be gained from experience using the brain's self-organizing approach. Thus we would have to use this type of experience-based approach. Nevertheless, the result you describe could be achieved. We could even include in these "loaded" (or learned) "skills and memories," the memory of having been the original person who was cryonically suspended, including having made the decision to be suspended, having become ill, and so on.

False reanimation

And this process would indeed appear to be a successful reanimation. The doctors would point to the "reanimated' patient as the proof in the pudding. Interviews of this patient would reveal that he was very happy with the process, delighted that he made the decision to be cryonically suspended, grateful to Alcor and the doctors for their successful reanimation of him, and so on.

But this would be a false reanimation. This is clearly not the same person that was suspended. His "memories" of having made the decision to be suspended four or five decades earlier would be false memories. Given the technology available at this time, it would be feasible to create entirely new humans from a genetic code and an experience / learning loading program (which simulates the learning in a much higher speed substrate to create a design for the new person). So creating a new person would not be unusual. So all this process has accomplished is to create an entirely new person who happens to share the genetic code with the person who was originally suspended. It's not the same person.

One might ask, "Who cares?" Well no one would care except for the originally suspended person. And he, after all, is not around to care. But as we look to cryonic suspension as a means towards providing a "second chance," we should care now about this potential scenario.

It brings up an issue which I have been concerned with, which is "false" reanimations.

Now one could even raise this issue (of a false reanimation) if the reanimated person does have the exact CNP of the original. One could take the philosophical position that this is still a different person. An argument for that is that once this technology is feasible, you could scan my CNP (perhaps while I'm sleeping) and create a CNP-identical copy of me. If you then come to me in the morning and say "good news, Ray, we successfully created your precise CNP-exact copy, we won't be needing your old body and brain anymore," I may beg to differ. I would wish the new Ray well, but feel that he's a different person. After all, I would still be here.

So even if I'm not still here, by the force of this thought experiment, he's still a different person. As you and I discussed at the reception, if we are using the preserved person as a data repository, then it would be feasible to create more than one "reanimated" person. If they can't all be the original person, then perhaps none of them are.

However, you might say that this argument is a subtle philosophical one, and that, after all, our actual particles are changing all the time anyway. But the scenario you described of creating a new person with the same genetic code, but with a very different CNP created through a learning simulation, is not just a matter of a subtle philosophical argument. This is clearly a different person. We have examples of this today in the case of identical twins. No one would say to an identical twin, "we don't need you anymore because, after all, we still have your twin."

I would regard this scenario of a "false" reanimation as one of the potential failure modes of cryonics.

Reverse-engineering the brain

Finally, on the issue of the LTM (long term memory), I think this is a good point and an interesting perspective. I agree that an efficient implementation of the knowledge in a human brain (and I am referring here to knowledge in the broadest sense as not just classical long term memory, but all of our skills and competencies) would be far more compact that the 10^{16} bytes I have estimated for its actual implementation.

As we understand biological mechanisms in a variety of domains, we find that we can redesign them (as we reverse engineer their functionality) with about 10^6 greater efficiency. Although biological evolution was remarkable in its ingenuity, it did get stuck in particular paradigms.

It's actually not permanently stuck in that its method of getting unstuck is to have one of its products, homo sapiens, discover and redesign these mechanisms.

We can point to several good examples of this comparison of our human engineered mechanisms to biological ones. One good example is Rob Freitas' design for robotic blood cells, which are many orders of magnitude more efficient than their biological counterparts.

Another example is the reverse engineering of the human auditory system by Lloyd Watts and his colleagues. They have found that implementing the algorithms in software from the reverse engineering of specific brain regions requires about a factor of 10^6 less computation than the theoretical potential of the brain regions being emulated.

Another good example is the extraordinarily slow computing speed of the interneuronal connections, which have about a 5 millisecond reset time. Today's conventional electronic circuits are already 100 million (10^8) times faster. Three-dimensional molecular circuits (e.g., nanotube-based circuitry) would be at least 10^9 times faster. Thus if we built a human brain equivalent with the same number of simulated neurons and connections (not just simulating the human brain with a smaller number of units that are operating at higher speeds), the resulting nanotube-based brain would operate at least 10^9 times faster than its biological counterpart.

Some of the inefficiency of the encoding of information in the human brain has a positive utility in that memory appears to have some holographic properties (meaningful information being distributed through a region), and this helps protect the information. It explains the usually gradual (as opposed to catastrophic) degradation of human memory and skill. But most of the inefficiency is not useful holographic encoding, but just this inherent inefficiency of biological mechanisms. My own estimate of this factor is around 10^6, which would reduce the LTM from my estimate of 10^{16} for the actual implementation to around 10^{10} for an efficient representation, but that is close enough to your and Minsky's estimate of 10^9.

However, as you point out, we don't know the compression/decompression algorithm, and are not in any event preserving this efficient representation of the LTM with the suspended patients. So we do need to preserve the inefficient representation.

With deep appreciation for your own contributions.

Eric Drexler: With respect to inferring memory state, the neurotransmitter-handling machinery in a synapse differs profoundly from the circuit structure in a DRAM cell. Memory cells in a chip are all functionally identical, each able to store and report different data from millisecond to millisecond; synapses in a brain are structurally diverse, and their differences encode relatively stable information. Charge stored in a DRAM cell varies without changes in its stable structure; long-term neurotransmitter levels in a synapse vary as a result of changes in its stable structure. The quantities of different enzymes, transport molecules, and so forth, determine the neurotransmitter properties relevant to LTM, hence neurotransmitter levels per se needn't be preserved.

My discussion of the apparent information-theoretic size of human LTM wasn't intended to suggest that such a compressed representation can or should be extracted from the detailed data describing brain structures. I expect that any restoration process will work with these far larger and more detailed data sets, without any great degree of intermediate compression. Nonetheless,

the apparently huge gap between the essential mental information to be preserved and the vastly more detailed structural information is reassuring—and suggests that *false* reanimation, while possible, shouldn't be expected when suspension occurs under good conditions. (Current medical practice has analogous problems of false life-saving, but these don't define the field.)

Ray Kurzweil: I'd like to thank you for an engaging dialogue. I think we've converged to a pretty close common vision of these future scenarios. Your point is well taken that human memory (for all of its purposes), to the extent that it involves the neurotransmitters, is likely to be redundantly encoded. I agree that differences in the levels of certain molecules are likely to be also reflected in other differences, including structural differences. Most biological mechanisms that we do understand tend to have redundant information storage (although not all; some single-bit changes in the DNA can be catastrophic). I would point out, however, that we don't yet understand the synaptic structures sufficiently to be fully confident that the differences in neurotransmitter levels that we need (for reanimation) are all redundantly indicated by structural changes. However, all of this can be tested with today's technology, and I would suggest that this would be worthwhile.

I also agree that "the apparently huge gap between the essential mental information to be preserved and the vastly more detailed structural information is reassuring." This is one example in which the inefficiency of biology is helpful.

Eric Drexler: Thank you, Ray. I agree that we've found good agreement, and I also enjoyed the interchange.

Additional comments on Jan. 15, 2003 by Robert Bradbury

Robert Bradbury: First, it is reasonable to assume that within this decade we will know the precise crystal structure for all human proteins for which cryonics reanimation is feasible, using either X-ray, NMR or computational (e.g., Blue Gene) based methods. That should be almost all human proteins. Second, it seems likely that we will have both the experimental (yeast 2-hybrid) or computational (Blue Gene and extensions thereof and/or distributed protein modeling, via @Home) to determine how proteins that interact typically do so. So we will have the ability to completely understand what happens at synapses and to some extent model that computationally.

Now, Ray placed an emphasis on neurotransmitter "concentration" that Eric seemed to downplay. I tend to lean in Eric's direction here. I don't think the molecular concentration of specific neurotransmitters within a synapse is particularly critical for reanimating a brain. I *do* think the concentrations of the macroscale elements necessary for neurotransmitter release will need to be known. That is, one needs to be able to count mitochondria and synaptic vesicle size and type (contents) as well as the post-synaptic neurotransmitter receptors and the pre-synaptic reuptake receptors. It is the numbers of these "machines of transmission" that determines the Hebbian "weight" for each synapse, which is a point I think Ray was trying to make.

Furthermore, if there is some diffusion of neurotransmitters out of individual synapses, the location and density of nearby synapses may be important (see Rusakov & Kullmann below). Now, the counting of and determination of the location of these "macroscale" effectors of

synapse activity is a much easier task than measuring the concentration of every neurotransmitter molecule in the synaptic cleft.

The neurotransmitter concentration may determine the instantaneous activity of the synapse, but I do not believe it holds the "weight" that Ray felt was important. That seems to be contained much more in the energy resources, enzymatic manufacturing capacity, and vesicle/receptor concentrations, which vary over much longer time periods. (The proteins have to be manufactured near the neuronal nucleus and be transported, relatively slowly, down to the terminal positions in the axons and dendrites.)

One can alter neurotransmitter concentrations and probably pulse-transmission probabilities at least within some range without disrupting the network terribly (risking false reanimation). SSRIs [Selective Serotonin Reuptake Inhibitors] and drugs used to treat Parkinson's, such as L-dopa, are examples of drugs that may alter these aspects of interneuron communications. Of more concern to me is whether or not there will be hurdles in attempting a "cold" brain restart. One can compare this to the difficulties of restarting the brain of someone in a coma and/or someone who has drowned.

The structure of the brain may be largely preserved but one just may not be able to get it running again. This implies there is some state information contained within the normal level of background activity. We haven't figured out yet how to "shock" the brain back into a functional pattern of activity.

Ray also mentioned vitrification. I know this is a hot topic within the cryonics community because of Greg Fahy's efforts. But you have to realize that Greg is trying to get us to the point where we can preserve organs entirely without nanotech capabilities. I think vitrification is a red herring. Why? Because we will know the structure of just about everything in the brain under 50 nm in size. Once frozen, those structures do not change their structure or location significantly.

So I would argue that you could take a frozen head, drop it on the floor so it shatters into millions or billions of pieces and as long as it remains frozen, still successfully reassemble it (or scan it into an upload). In its disassembled state it is certainly one very large 3D jigsaw puzzle, but it can only be reassembled one correct way. Provided you have sufficient scanning and computational capacity, it shouldn't be too difficult to figure out how to put it back together.

You have to keep in mind that all of the synapses have proteins binding the pre-synaptic side to the post-synaptic side (e.g., molecular velcro). The positions of those proteins on the synaptic surfaces are not specified at the genetic level and it seems unlikely that their locations would shift significantly during the freezing process (such that their number and approximate location could not be reconstructed).

As a result, each synapse should have a "molecular fingerprint" as to which pre-side goes with which post-side. So even if the freezing process pulls the synapse apart, it should be possible to reconstruct who the partners are. One needs to sit and study some freeze-fracture electron micrographs before this begins to become a clear idea for consideration.

So I think the essential components are the network configuration itself, the macroscale machinery architecture of the synapses and something that was not mentioned, the "transcriptional state of the nuclei of the neurons" (and perhaps glial cells), i.e., which genes are turned on/off. This may not be crucial for an instantaneous brain "reboot" but might be essential for having it function for more than brief periods (hours to days).

References

A good (relatively short but detailed) description of synapses and synaptic activity is Ch.5: Synaptic Activity from State University of New York at Albany.

Also see:

Understanding Neurological Functions through the Behavior of Molecules, Dr. Ryoji Yano

Three-Dimensional Structure of Synapses in the Brain and on the Web, J. C. Fiala, 2002 World Congress on Computational Intelligence, May 12-17, 2002

Assessing Accurate Sizes of Synaptic Vesicles in Nerve Terminals, Seongjai Kim, Harold L. Atwood & Robin L. Cooper

Extrasynaptic Glutamate Diffusion in the Hippocampus: Ultrastructural Constraints Uptake and Receptor Activation, Dimitri A. Rusakov & Dimitry M. Kullmann The J. of Neuroscience 18(9):3158-3170 (1 May 1998).

Ray Kurzweil: Robert, thanks for your interesting and thoughtful comments. I essentially agree with what you're saying, albeit we don't yet understand the mechanisms behind the "Hebbian weight" or other vital state information needed for a non-false reanimation. It would be good if this state information were fully represented by mitochondria and synaptic vesicle size and type (contents), post-synaptic neurotransmitter receptors and pre-synaptic reuptake receptors, i.e., by the number of these relatively large (compared to molecules) "machines of transmission."

Given that we have not yet reverse-engineered these mechanisms, I suppose it would be difficult to do a definitive experiment now to make sure we are preserving the requisite information.

I agree with your confidence that we will have reverse-engineered these mechanisms within the next one to two decades. I also agree that we need only preserve the information, and that reanimation technology will take full advantage of the knowledge of how these mechanisms work. Therefore the mechanisms don't need to preserved in working order so long as the information is there. I agree that Fahy's concerns apply primarily to revitalization without such detailed nanotech repair and reconstruction.

Of course, as I pointed out in the debate with Eric, such a complete reconstruction may essentially amount to creating a new brain/person with the cryonically preserved brain/body serving only as a blueprint, in which case it would just as easy to create more than one renaimated person. Eric responded to this notion by saying that the first one is the reanimated

person and subsequent ones are just copies because after all, at that time, we could make copies of anyone anyway.

With regard to your jigsaw puzzle, that may be a difficult puzzle to put together, although I suppose we'll have the computational horsepower to do it.

The Alcor Conference on Extreme Life Extension

Ray Kurzweil

http://www.kurzweilai.net/articles/art0531.html

On November 15-17, 2002, leaders in life extension and cryonics came together to explore how the emerging technologies of biotechnology, nanotechnology, and cryonics will enable humans to halt and ultimately reverse aging and disease and live indefinitely.

Published on KurzweilAI.net November 22, 2002. Additional reporting by Sarah Black.

The idea that death is inevitable, which I call the "death meme," is a powerful and pervasive belief held by all humans, with the exception of a small but growing group of life extensionists. The thought leaders of this movement gathered together this past weekend in Los Angeles to participate in the Fifth annual Alcor Conference on Extreme Life Extension and share ideas on pushing back the end of life. Bringing together longevity experts, biotechnology pioneers, and futurists, the conference explored how the emerging technologies of biotechnology, nanotechnology, and cryonics will enable humans to halt and ultimately reverse aging and disease and live indefinitely.

I had the opportunity to participate in this illuminating and stimulating conference and I report herein on the highlights.

Robert Freitas is a Research Scientist at Zyvex, a nanotechnology company, and in my view the world's leading pioneer in nanomedicine. He is the author of a book by the same name and the inventor of a number of brilliant conceptual designs for medical nanorobots. In his first major presentation of his pioneering conceptual designs, Freitas began his lecture by lamenting that "natural death is the greatest human catastrophe." The tragedy of medically preventable natural deaths "imposes terrible costs on humanity, including the destruction of vast quantities of human knowledge and human capital." He predicted that "future medical technologies, especially nanomedicine, may permit us first to arrest, and later to reverse, the biological effects of aging and most of the current causes of natural death."

Freitas presented his design for "respirocytes," nanoengineered replacements for red blood cells. Although they are much smaller than biological red blood cells, an analysis of their functionality demonstrates that augmenting one's blood supply with these high pressure devices would enable a person to sit at the bottom of a pool for four hours, or to perform an Olympic sprint for 12 minutes, without taking a breath. Freitas presented a more complex blueprint for robotic "microbivores," white blood cell replacements that would be hundreds of times faster than normal white blood cells.

By downloading appropriately updated software from the Internet, these devices would be quickly effective against any type of pathogen, including bacteria, viruses, fungi, and cancer cells. Freitas also presented a new concept of a "chromosome replacement robot," which would be programmed to enter a cell nucleus and perform repairs and modifications to a person's DNA

to reverse DNA transcription errors and reprogram defective genetic information. Trillions of such robots could be programmed to enter every cell in the body.

How we will get to this kind of technology was the subject of my [**Ray Kurzweil**] presentation on the law of accelerating returns at the conference. Communication bandwidths, the shrinking size of technology, our knowledge of the human brain, and human knowledge in general are all accelerating. Three-dimensional molecular computing will provide the hardware for human-level "strong" AI well before 2030. The more important software insights will be gained in part from the reverse-engineering of the human brain, a process well under way. The ongoing acceleration of price-performance of computation, communication, and miniaturization will provide the technologies to create nanobots that can instrument (place sensors in) billions of neurons and interneuronal connections, greatly facilitating the development of detailed models of how human intelligence works.

Once nonbiological intelligence matches the range and subtlety of human intelligence, it will necessarily soar past it because of the continuing acceleration of information-based technologies, as well as the ability of machines to instantly share their knowledge. Intelligent nanorobots will be deeply integrated in the environment, our bodies and our brains, providing vastly extended longevity, full-immersion virtual reality incorporating all of the senses, experience "beaming," and enhanced human intelligence. The implication will be an intimate merger between the technology-creating species and the evolutionary process it spawned.

Aubrey de Grey, a researcher at the University of Cambridge, began his talk by citing the fact that 100,000 people die of age-related causes each day, and then quoted Bertrand Russell's statement that "some of us think this is rather a pity." (Albeit Russell was talking about nuclear war rather than aging.) de Grey described a program he has devised to approach the goal of extreme life extension "with a hard-headed, engineering frame of mind." He described his goal as "engineered negligible senescence," referring to the term "negligible senescence" that Tuck Finch introduced in 1990, defined as "the absence of a statistically detectable inverse correlation between age and remaining life expectancy."

Human society takes for granted the existence of this inverse correlation (between age and remaining life expectancy), but de Grey explained why he feels we have the knowledge close at hand to flatten out this curve. His program (to develop engineered negligible senescence) "focuses mainly on those subtle changes, the ones that accumulate throughout life and only snowball into pathology rather late. That's why 'engineered negligible senescence' is an accurate term for my goal—I aim to eliminate those subtle changes, so allowing the cell/organ/body to use its existing homeostatic prowess to maintain us in a physically un-deteriorating state indefinitely."

de Grey argued persuasively for the feasibility of this goal and described a multi-faceted program to address each known area of aging, including his area of specialty in mitochondrial mutations and lysosomal aggregates. He proposed an "Institute of Biomedical Gerontology," with a budget of $100 million, to promote, coordinate, and fund the focused development of these rejuvenation biotechnologies.

Christine Peterson, cofounder and President of the Foresight Institute, provided guidelines on how the lay person can evaluate the often conflicting advice and information on health and life extension. Christine pointed out that as knowledge becomes increasingly specialized, no one person can be an expert in every treatment intervention, so "we are all lay persons" even if we have expertise in some particular aspect of health treatment. She pointed out the destructive implications of the benign sounding creed of the medical profession, "first of all, do no harm." Because of the extremely cautious, risk-adverse orientation that this principle fosters, treatments desperately needed by millions of people are tragically suppressed or delayed.

Max More, President of the Extropy Institute, and the Futures specialist at ManyWorlds, Inc., presented what he called a "strategic scenario analysis for your second life." More described his own culture shock at having moved from England to Southern California, which led him to consider the extreme adjustment challenge for people (possibly himself) in the future being reanimated from cryonic suspension. More pointed out that "to maximize our chances of a psychologically successful revival, we have the responsibility to prepare ahead of time." Using the discipline of scenario thinking from his consulting work, More engaged in a series of thought experiments that he would encourage people to engage in who have made the decision to be cryonically suspended should they happen to die.

Michael West, President and CEO of Advanced Cell Technology, Inc. and a pioneer of therapeutic cloning, presented a compelling history of the science of cellular aging. He emphasized the remarkable stability of the immortal germ line cells, which link all cell-based life on Earth. He described the role of the telomeres, a repeating code at the end of each DNA strand, which are made shorter each time a cell divides, thereby placing a limit on the number of times a cell can replicate (the "Hayflick limit"). Once these DNA "beads" run out, a cell becomes programmed for cell death. The immortal germ line cells avoid this destruction through the use of a single enzyme called telomerase, which rebuilds the telomere chain after each cell division. This single enzyme makes the germ line cells immortal, and indeed these cells have survived from the beginning of life on Earth billions of years ago.

This insight opens up the possibility of future gene therapies that would return cells to their youthful, telomerase-extended state. Animal experiments have shown telomerase to be relatively benign, although some experiments have resulted in increased cancer rates. There are also challenges in transferring telomerase into the cell nuclei, although the gene therapy technology required is making solid progress. West expressed confidence that new techniques would provide the ability to transfer the telomerase into the nuclei, and to overcome the cancer issue. Telomerase gene therapy holds the promise of indefinitely rejuvenating human somatic (non-germ line) cells i.e., all human cells.

West addressed the ethical controversies surrounding stem cell therapies. He pointed out a number of inconsistencies in the ethical position of those who oppose stem cell therapies. For example, a fetus can divide in two, within the first two weeks after conception and prior to implantation in the mother's womb, to create identical twins. This demonstrates that a unique human life is not defined by a fertilized egg cell, but only by an implanted embryo. Stem cell therapies use fetal cells prior to this individuation process. West pointed out the dramatic health benefits that stem cell therapies promise, including the ability to create new cells and organs to

treat a wide variety of diseases such as Parkinson's disease and heart disease. West also described promising new methodologies in the field of "human somatic cell engineering" to create new tissues with a patient's own DNA by modifying one type of cell (such as a skin cell) directly into another (such as a pancreatic Islet cell or a heart cell) without the use of fetal stem cells.

Greg Fahy, Chief Scientific Officer of 21st Century Medicine, formerly director of an organ cryopreservation program at the American Red Cross and a similar program for the Naval Medical Research Institute, described prospects for preserving organs for long periods of time. He pointed out how we now have "the ability to perfuse whole kidneys with cryoprotectants at concentrations that formerly were uniformly fatal, but which currently produce little or no injury."

The immediate goal of Fahy's research is to preserve transplant organs for substantially longer periods of time than is currently feasible. Fahy pointed out that by combining these techniques with the therapeutic cloning technologies being developed by Michael West and his colleagues, it will be possible in the future for people to keep a supply of replacements for all of their organs, to be immediately available in emergencies. He painted a picture "of the future when organs are grown, stored, and transported as easily as blood is today."

To suggest a way to make it to that future, I [**Ray Kurzweil**] had the opportunity to present a set of ideas to apply our current knowledge to life extension. My earlier presentation focused on the nature of human life in the 21st century, whereas this presentation described how we could live to see (and enjoy!) the century ahead. These ideas are drawn from an upcoming book, A Short Guide to a Long Life, which I am coauthoring with Terry Grossman, M.D., a leading longevity expert.

These ideas should be thought of as "a bridge to a bridge to a bridge," in that they provide the means to remain healthy and vital until the full flowering of the biotechnology revolution within 20 years, which in turn will bring us to the nanotechnology-AI (artificial intelligence) revolution ten years after that. The latter revolution will radically redefine our concept of human mortality.

I pointed out that the leading causes of death (heart disease, cancer, stroke, diabetes, kidney disease, liver disease) do not appear out of the blue. They are the end result of processes that are decades in the making. You can understand where you are personally in the progression of these processes and end (and reverse) the lethal march towards these diseases. The program that Dr. Grossman and I have devised allows you to assess how longstanding imbalances in your metabolic processes can be corrected before you "fall off the cliff." This information is not "plug and play," but the knowledge is available and can be applied through a comprehensive and concerted effort.

The nutritional program that Dr. Grossman and I recommend provides the best of the two contemporary poles of nutritional thinking. The Atkins philosophy has correctly identified the dangers of a high-glycemic-index diet as causing imbalances in the sugar and insulin cycle, but does not focus on the equally important rebalancing of omega 3 and omega 6 fats, and cutting down on the pro-inflammatory fats in animal products. Conversely, the low-fat philosophy of

Ornish and Pritikin has not placed sufficient attention on cutting down on high-glycemic-index starches. Our program recommends a moderately low level of carbohydrates, dramatic reductions in high-glycemic-index carbohydrates, as well as moderately low levels of fat, with an emphasis on the anti-inflammatory Omega-3 fats found in nuts, fish, and flaxseed.

A study of nurses showed that those nurses who ate at least a handful of nuts (one ounce) each day had 75% less heart disease than the nurses who did not eat nuts. Our program also includes aggressive supplementation to obtain optimal lipid levels, reduce inflammation, correct potential problems with the methylation (folic acid) cycle, attain and maintain an optimal weight, and maintain glucose and insulin levels in a healthy balance.

In a rare lecture, **Eric Drexler**, author of Engines of Creation, the seminal book that introduced the field two decades ago, and widely regarded as the father of nanotechnology, reflected on the state of the nanotechnology field and its prospects. Drexler pointed out that the term "nanotechnology" has broadened from his original conception, which was the precise positional control of chemical reactions to any technology that deals with measurements of less than 100 nanometers. Drexler pointed to biology as an existence proof of the feasibility of molecular machines. Our human-designed machines, Drexler pointed out, will not be restricted to the limitations of biology. He said that although the field was initially controversial, no sound criticism has emerged for his original ideas. Drexler dramatically stated, "I therefore declare victory by default."

Drexler cited the powerful analogy relating atoms and bits to nanotechnology and software. We can write a piece of software to perform a certain manipulation on several numbers. We can then use logic and loops to perform that same manipulation billions or trillions of times, even though we only have to write the software once. Similarly, we can set up nanotechnology systems to perform the same nanoscale mechanical manipulations billions or trillions of times and in billions or trillions of locations.

Drexler described the broad applicability of nanotechnology to revolutionize many areas of human endeavor. We will be able to build supercomputers that are one thousandth of the size of a human cell. We will be able to create electricity-generating solar panels at almost no cost. We will be able to build extremely inexpensive spacecraft out of diamond fiber. "The idea that our human world is limited to the Earth is going to be obsolete very soon, as soon as these technologies become available," Drexler pointed out. Indeed, all manufacturing will be revolutionized. Nanotechnology-based manufacturing will make feasible the ability to create any customized product we can define at extremely low cost from inexpensive raw materials and software.

With regard to our health, nanotechnology will be able to reconstruct and rebuild just about everything in our bodies. Nanoscale machines will enter all of our cells and proofread our DNA, patch the mitochondria, destroy pathogens, remove waste materials, and rebuild our bodies and brains in ways unimaginable today. Drexler defined this goal as "permanent health."

Drexler expressed optimism for the prospects of successful reanimation of cryonically preserved people. Nanorobots will be able to assess, analyze, and investigate the state of the preserved

cells, tissues, and fluids; perform microscopic and nanoscopic repairs on every cell and connection, and remove cryopreservatives. He chided other cryonics supporters for making the "pessimistic argument" that although cryonics had only a small chance of working, this chance was better than the alternative, which provided no chance for a second life. Based on our growing knowledge and confidence in nanotechnology and emerging scenarios for applying these technologies to the reanimation task, Drexler argued that we should be expressing a valid optimism about the prospects for a healthy second life after suspension.

Drexler was asked what he thought of the prospects for optical and quantum computing. He replied that optical computers will remain bulkier than programmable molecular computers and thus are likely to remain special purpose devices. As for quantum computing, there are designs for possible room-temperature quantum computers with dozens of qubits, but the prospects for quantum computing are still not clear.

Drexler was pessimistic on the prospects for picotechnology (technology on a scale 1000 times smaller than nanotechnology). He explained that one would need the conditions of a neutron star to make this feasible, and even then there are theoretical problems getting subatomic particles to perform useful functions such as computation.

I would point out that nanotechnology also appeared unlikely until Drexler came along and showed how we could build machines that go beyond the nanomachines of nature. A future Drexler is likely to provide the conceptual designs to build machines that go beyond the picomachines of atomic nuclei and atoms.

I have that penciled in for 2072.

Arguments for a Green and Gray Future

Ray Kurzweil

http://www.kurzweilai.net/articles/art0462.html

Ray Kurzweil and Gregory Stock, Director, UCLA Program on Medicine, Technology and Society, debated "BioFuture vs. MachineFuture" at the Foresight Senior Associate Gathering, April 27, 2002. This is Ray Kurzweil's presentation.

Published on KurzweilAI.net May 1, 2002.

Audio clips of the debate

The Future Will be Both Green AND Gray

The First 2 Decades of the 21st Century will be the Golden Age of Biotechnology

- We've reached the intersection of Biology and Information Science
    - Biology as software
    - We're learning the information transformations underlying life and disease processes
        - How they work
        - How to fix / transform / enhance these natural methods

Many Intersecting Bio-Information Revolutions

- Tissue engineering: grow new telemere-extended cells, tissues, organs
- Rational drug design: design drugs for precision tasks
- Genomic panels
- Fixing genomic defects
- Reverse-engineering the Genome through the Proteome
    - Precise tracking of each individual's biochemical pathways
- Individualized medicine
- And many others. . . . .

The 2nd, 3rd, and 4th Decades will be the Golden Age of BioNanoTech

- We've already crossed the threshold:
    - We have devices emerging to replace body parts, organs
    - The age of neural implants is well under way
        - Cochlear implants
        - Parkinson's Implant (communicates directly with ventral posterior nucleus)

        - Experimental implants for stroke patients
        - Retinal implants
        - Many others under development
    - There are already 4 major conferences on BioMEMS

Many Emerging Designs for Linking the Wet Analog World of Biological Information with Electronic Information

- Max Planck Institute's Noninvasive "Neuron Transistor"
- Quantum Dots
- Control of Neuroprosthetic systems using pattern recognition on brain activity despite damaged nerve pathway

Intelligent Machines are Making Their Way Into Our Blood Stream

- U of Illinois at Chicago capsules with 7 nanometer pores cured type I Diabetes in Rats
- Many designs to deliver medications in controlled manner, including the brain
- Sandia micro robot traps cells with tiny jaws and implants substances
- Robert Freitas' conceptual designs for respirocytes, artificial platelets, nanorobotic microbivores
- Many other examples….

Nanotech is behind Biotech, but….

- Consider the law of accelerating returns
- We have been and will continue to double the paradigm shift rate each decade
    - So the 20th century was < 20 years of progress at **today's** rate of progress
    - We'll make equivalent progress in < 15 years
    - The 21st century is equivalent to 20,000 years of progress at **today's** rate of progress

By 2030

- Electronics will utilize molecule-sized circuits….
- organized in three dimensions
- bioMEMS will have evolved into bioNEMS

A Big Role for Small Robots

- It will be routine to have millions / billions of nanobots coursing through our capillaries, communicating with:

    - Each other via a wireless LAN
    - The Internet
    - Our biological neurons
- Providing:
    - Vast augmentation to our immune system
    - Otherwise repairing and augmenting our biological systems
    - Providing full immersion virtual reality encompassing all 5 senses
        - and neurological correlates of our emotions
    - "Experience Beaming"
    - Most importantly…..

Expanding our Minds…

- Multiplying our mere hundred trillion connections many fold (eventually by trillions)
- Intimate connection with nonbiological forms of intelligence
- Direct connection to other minds
- Downloading of knowledge

Nonbiological Intelligence will combine….

- The parallel self-organizing paradigm of biological pattern recognition, with
- The strengths of machine intelligence:
    - Speed
    - Accuracy and scale of memory
    - Ability to instantly share knowledge
    - Ability to pool and network resources

The Ethical Barriers are very weak

- The ethical barriers even for biological technology are weak:
    - Like stones in a stream, the water rushes around them
        - e.g., the stem cell controversy has only accelerated efforts to bypass unneeded egg cells by transforming one cell type into another
            - through an understanding of the protein signaling factors

"Natural" Technologies Always Proceed Synthetic Technologies

- Carrier pigeons were eclipsed by human made flying machines · Human scribes were replaced
- Human scribes were replaced by automated word processing
- Machines greatly outperform human and animal labor

Ultimately AI will vastly outperform human intelligence

- Biological intelligence is stuck
  - Biological humanity (10^{10} humans) has 10^{26} calculations per second today and will have 10^{26} cps 50 years from now
- Machine intelligence is millions of times less powerful today, but growing at a double exponential rate
- The cross over point is in the 2020s
- By 2050, nonbiological intelligence will be trillions of times more powerful than biological intelligence
- Machine intelligence will combine the strengths of both contemporary human intelligence with the speed, capacity, and knowledge sharing of machines

The perspective that this "Singularity" in human history is a century or more away fails to appreciate the explosive nature of the exponential growth inherent in the law of accelerating returns.

Gregory Stock's presentation

Ray Kurzweil's presentation

Debate

Audience Q&A

Foreword to ‘Dark Ages II’

Ray Kurzweil

(Book By Bryan Bergeron)

http://www.kurzweilai.net/articles/art0227.html

Our civilization’s knowledge legacy is at great risk, growing exponentially with the exploding size of our knowledge bases. And that doesn’t count the trillions of bytes of information stored in our brains, which eventually will be captured in the future. How long do we want our lives and thought to last?

Foreword published on KurzweilAI.net July 26, 2001. Book published by Prentice Hall September 28, 2001.

My father was one of those people who liked to store all the images and sounds that documented his life. So upon his untimely death at the age of 58 in 1970, I inherited his archives which I treasure to this day. I have my father's 1938 doctoral dissertation at the University of Vienna containing his unique insights into the contributions of Brahms to our musical vocabulary. There are albums of neatly arranged newspaper clippings of his acclaimed musical concerts as a teenager in the hills of Austria. There are the urgent letters to and from the American music patrons who sponsored his flight from Hitler just before "Krystalnacht" made such escape impossible. These items are among dozens of aging boxes containing a myriad of old remembrances, including photographs, musical recordings on vinyl and magnetic tape, personal letters, and even old bills.

I also inherited his penchant for preserving the records of one's life, so along with my father's boxes, I have several hundred boxes of my own. My father's productivity assisted by the technology of his manual typewriter and carbon paper cannot compare with my own prolificacy, aided and abetted by computers and high speed printers which can reproduce my thoughts in all kinds of permutations.

Tucked away in my own boxes are also various forms of digital media: punch cards, paper tape reels, and digital magnetic tapes and disks of various sizes and formats. I often think about just how accessible this information remains. Ironically, the ease of approaching this information is inversely proportional to the level of advancement of the technology used to create it. Most straightforward are the paper documents, which although showing the signs of age, are imminently readable. Only slightly more challenging are the vinyl records and analog sound tape recordings. Although some basic equipment is required, these are not difficult items of equipment to find or use. The punch cards are somewhat more difficult, but it's still possible to find punch card readers, and the formats are uncomplicated.

By far, the most difficult information to retrieve is that contained on the digital disks and tapes. Consider the challenges involved. For each one, I have to figure out exactly which disk or tape drive was used. I then have to recreate the exact hardware configuration from many years ago.

Try finding an IBM 1620 circa 1960 or Data General Nova I circa 1973 with exactly the right disk drive and controller, and you'll quickly discover the difficulties involved. Then once you've assembled the requisite old equipment, there are layers of software to deal with: the appropriate operating system, disk information drivers, and application programs. Then just who are you going to call when you run into the inevitable scores of problems inherent in each layer of hardware and software? It's hard enough getting contemporary systems to work, let alone systems for which the help desks were disbanded decades ago. Even the Computer Museum, which used to be located in Boston, has been disbanded, and even when it was in business, most of the old computers on display had stopped functioning many years earlier.

Assuming that you prevail through all of these obstacles, the actual magnetic data on the disks has probably decayed. So even if we assume that the old hardware and software that you assembled are working perfectly, and that you have aging human experts to assist you with perfect recall of long since obsolete equipment, these old computers would still generate mostly error messages.

So is the information gone? The answer is: not entirely. Even though the magnetic spots may no longer be readable by the original equipment, the faded magnetic regions could be enhanced by suitably sensitive equipment using methods that are analogous to the image enhancement often used on images of the pages of old books. So the information is still there, albeit extremely difficult to get at. With enough devotion and historical research one might actually retrieve it. If we had reason to believe that one of these disks contained secrets of enormous value, we would probably succeed in recovering the information. But the mere motivation of nostalgia is unlikely to be sufficient for this formidable task. I will say that I did largely anticipate this problem, so I do have paper print outs of most of these old files. Invariably, that will be how I solve this problem. The bottom line is that accessing information stored in digital form decades (and sometimes even just years) later is extremely difficult if not impossible.

However, keeping all our information on paper is not the answer. Hard copy archives present a different problem. Although I can readily read even a century-old paper manuscript if I'm holding it in my hand, finding a desired document from among thousands of only modestly organized file folders can be a frustrating and time consuming task. It can take an entire afternoon to locate the right folder, not to mention the risk of straining one's back from moving dozens of heavy file boxes from one stack to another. Using the more compact form of hard copy known as microfilm or microfiche may alleviate some of the problems, but the difficulties of locating the right document remain.

So I have had a dream of taking all of these archives, scanning them into a massive personal data base, and then being able to utilize powerful contemporary search and retrieve methods on the hundreds of thousands of scanned and OCR'd (Optical Character Recognized) records. I even have a name for this project: DAISI (Document And Image Storage Invention), and I have been accumulating the ideas for this little venture for many years.

DAISI will involve the rather formidable task of scanning and OCR'ing hundreds of thousands of documents, and patiently cataloguing them into a data base. But the real challenge to my dream of DAISI is the one that Bryan Bergeron articulates so eloquently in this volume, namely how

can I possibly select appropriate hardware and software layers that will give me the confidence that my archives will be viable and accessible decades from now?

Of course my own archival desires are a microcosm of the exponentially expanding knowledge base that the human civilization is accumulating. It is this shared species-wide knowledge base that distinguishes us from other animals. Other animals communicate, but they don't accumulate an evolving and growing base of knowledge to pass down to the next generation. Given that we are writing our precious heritage in what Bergeron calls "disappearing ink," our civilization's legacy would appear to be at great risk. The danger appears to be growing exponentially along with the exploding size of our knowledge bases. The problem is further exacerbated by the accelerating speed with which we turn over to new standards in the many layers of hardware and software needed to store information.

Is there an answer to this dilemma? Bergeron's insightful volume articulates the full dimension of the problem as well as a road map to ameliorating its destructive effects. I will summarize my own response to this predicament below, but first we need to consider yet another source of knowledge.

There is another valuable repository of information stored in our own brains. Our memories and skills, although they may appear to be fleeting, do represent information, stored in vast patterns of neurotransmitter concentrations, interneuronal connections, and other salient neural details. I have estimated the size of this very personal data base at thousands of trillions of bytes (per human brain), and we are further along than many people realize in being able to access this data and understand its encoding. We have already "reverse engineered" (i.e., scanned and understood the methods of) several dozen of the hundreds of regions of the brain, including the way in which information is coded and transmitted from one region to another.

I believe it is a conservative scenario to say that within thirty years we will have completed the high resolution scan of the human brain (just as we have completed today the scan of the human genome) and will have detailed mathematical models of the hundreds of information processing organs we call the human brain. Ultimately we will be able to access and understand the thousands of trillions of bytes of information we have tucked away in each of our brains.

This will introduce the possibility of reinstantiating the vast patterns of information stored in our electrochemical neural mechanisms into other substrates (i.e., computational mechanisms) that will be much more capable in terms of speed, capacity, and in the ability to quickly share knowledge. Today, our brains are limited to a mere hundred trillion connections. Later in this century, our minds won't have to stay so small.

Copying our minds to other mediums raises some key philosophical issues, such as "is that really me," or rather someone else who just happens to have mastered all my thoughts and knowledge? Without addressing all of these issues in this foreword, I will mention that the idea of capturing the information and information processes in our brains has raised the specter that we (or at least entities that act very much like we do) could "live forever." But is that really the implication?

For eons, the longevity of our mental software has been inexorably linked to the survival of our biological hardware. Being able to capture and reinstantiate all the details of our information processes would indeed separate these two aspects of our mortality. But the profound implication of Bergeron's Dark Ages II is that software does not necessarily live forever. Indeed there are formidable challenges to it living very long at all.

So whether information represents one man's sentimental archive, or the accumulating knowledge base of the human-machine civilization, or the mind files stored in our brains, what can we say is the ultimate resolution regarding the longevity of software? The answer is simply this: information lasts only so long as someone cares about it. The conclusion that I've come to with regard to my DAISI project, after several decades of careful consideration, is that there is no set of hardware and software standards existing today, nor any likely to come along, that will provide me with any reasonable level of confidence that the stored information will still be accessible (without unreasonable levels of effort) decades from now. The only way that my archive (or any one else's) can remain viable is if it is continually upgraded and ported to the latest hardware and software standards. If an archive remains ignored, it will ultimately become as inaccessible as my old 8 inch disk platters.

In this pioneering work, Bergeron describes the full dimensions of this fundamental issue, and also provides a compelling set of recommendations to preserve key sources of information beyond the often short-sighted goals underlying the design of most contemporary information processing systems. The bottom line will remain that information will continue to require continual maintenance and support to remain "alive." Whether data or wisdom, information will only survive if we want it to.

We are continually recreating our civilization's trove of knowledge. It does not simply survive by itself. We are constantly rediscovering, reinterpreting, and reformatting the legacy of culture and technology that our forbears have bestowed to us. We will eventually be able to actually access the vast patterns of information in our brains, which will provide the opportunity to back up our memories and skills. But all of this information will be fleeting if no one cares about it. Translating our currently hardwired thoughts into software will not necessarily provide us with immortality. It will simply put the means to determine how long we want our lives and thoughts to last into our own figurative hands.

Dark Ages II

How to Build a Brain

A machine is likely to achieve the ability of a human brain in the coming years. Ray Kurzweil has predicted that a $1,000 personal computer will match the computing speed and capacity of the human brain by around the year 2020. With human brain reverse engineering, we should have the software insights before 2030. This section explores the possibilities of machine intelligence and exotic new technologies for faster and more powerful computational machines, from cellular automata and DNA molecules to quantum computing. It also examines the controversial area of uploading your mind into a computer.

The Intelligent Universe

Ray Kurzweil

http://www.kurzweilai.net/articles/art0534.html

Within 25 years, we'll reverse-engineer the brain and go on to develop superintelligence. Extrapolating the exponential growth of computational capacity (a factor of at least 1000 per decade), we'll expand inward to the fine forces, such as strings and quarks, and outward. Assuming we could overcome the speed of light limitation, within 300 years we would saturate the whole universe with our intelligence.

Published on KurzweilAI.net December 12, 2002. Originally published on http://www.edge.org November 7, 2002.

On July 21, 2002, Edge brought together leading thinkers to speak about their "universe." Other participants:

The Computational Universe by ***Seth Lloyd***
The Emotion Universe by ***Marvin Minsky***
The Inflationary Universe by *Alan Harvey Guth*
The Cyclic Universe by *Paul Steinhardt*

The universe has been set up in an exquisitely specific way so that evolution could produce the people that are sitting here today [at Edge's REBOOTING CIVILIZATION II meeting on July 21, 2002] and we could use our intelligence to talk about the universe. We see a formidable power in the ability to use our minds and the tools we've created to gather evidence, to use our inferential abilities to develop theories, to test the theories, and to understand the universe at increasingly precise levels. That's one role of intelligence. The theories that we heard on cosmology look at the evidence that exists in the world today to make inferences about what existed in the past so that we can develop models of how we got here.

Then, of course, we can run those models and project what might happen in the future. Even if it's a little more difficult to test the future theories, we can at least deduce, or induce, that certain phenomena that we see today are evidence of times past, such as radiation from billions of years ago. We can't really test what will happen billions or trillions of years from now quite as directly, but this line of inquiry is legitimate, in terms of understanding the past and the derivation of the universe. As we heard today, the question of the origin of the universe is certainly not resolved. There are competing theories, and at several times we've had theories that have broken down, once we acquired more precise evidence.

At the same time, however, we don't hear discussion about the role of intelligence in the future. According to common wisdom, intelligence is irrelevant to cosmological thinking. It is just a bit of froth dancing in and out of the crevices of the universe, and has no effect on our ultimate cosmological destiny. That's not my view. The universe has been set up exquisitely enough to have intelligence. There are intelligent entities like ourselves that can contemplate the universe

and develop models about it, which is interesting. Intelligence is, in fact, a powerful force and we can see that its power is going to grow not linearly but exponentially, and will ultimately be powerful enough to change the destiny of the universe.

I want to propose a case that intelligence—specifically human intelligence, but not necessarily biological human intelligence—will trump cosmology, or at least trump the dumb forces of cosmology. The forces that we heard discussed earlier don't have the qualities that we posit in intelligent decision-making. In the grand celestial machinery, forces deplete themselves at a certain point and other forces take over. Essentially you have a universe that's dominated by what I call dumb matter, because it's controlled by fairly simple mechanical processes.

Human civilization possesses a different type of force with a certain scope and a certain power. It's changing the shape and destiny of our planet. Consider, for example, asteroids and meteors. Small ones hit us on a fairly regular basis, but the big ones hit us every some tens of millions of years and have apparently had a big impact on the course of biological evolution. That's not going to happen again. If it happened next year we're not quite ready to deal with it, but it doesn't look like it's going to happen next year. When it does happen again our technology will be quite sufficient. We'll see it coming, and we will deal with it. We'll use our engineering to send up a probe and blast it out of the sky. You can score one for intelligence in terms of trumping the natural unintelligent forces of the universe.

Commanding our local area of the sky is, of course, very small on a cosmological scale, but intelligence can overrule these physical forces, not by literally repealing the natural laws, but by manipulating them in such a supremely sublime and subtle way that it effectively overrules these laws. This is particularly the case when you get machinery that can operate at nano and ultimately femto and pico scales. Whereas the laws of physics still apply, they're being manipulated now to create any outcome the intelligence of this civilization decides on.

How intelligence developed

Let me back up and talk about how intelligence came about. Wolfram's book has prompted a lot of talk recently on the computational substrate of the universe and on the universe as a computational entity. Earlier today, Seth Lloyd talked about the universe as a computer and its capacity for computation and memory. What Wolfram leaves out in talking about cellular automata is how you get intelligent entities. As you run these cellular automata, they create interesting pictures, but the interesting thing about cellular automata, which was shown long before Wolfram pointed it out, is that you can get apparently random behavior from deterministic processes.

It's more than apparent that you literally can't predict an outcome unless you can simulate the process. If the process under consideration is the whole universe, then presumably you can't simulate it unless you step outside the universe. But when Wolfram says that this explains the complexity we see in nature, it's leaving out one important step. As you run the cellular automata, you don't see the growth in complexity—at least, certainly he's never run them long enough to see any growth in what I would call complexity. You need evolution.

Marvin talked about some of the early stages of evolution. It starts out very slow, but then something with some power to sustain itself and to overcome other forces is created and has the power to self-replicate and preserve that structure. Evolution works by indirection. It creates a capability and then uses that capability to create the next. It took billions of years until this chaotic swirl of mass and energy created the information-processing, structural backbone of DNA, and then used that DNA to create the next stage.

With DNA, evolution had an information-processing machine to record its experiments and conduct experiments in a more orderly way. So the next stage, such as the Cambrian explosion, went a lot faster, taking only a few tens of millions of years. The Cambrian explosion then established body plans that became a mature technology, meaning that we didn't need to evolve body plans any more.

These designs worked well enough, so evolution could then concentrate on higher cortical function, establishing another level of mechanism in the organisms that could do information processing. At this point, animals developed brains and nervous systems that could process information, and then that evolved and continued to accelerate. Homo sapiens evolved in only hundreds of thousands of years, and then the cutting edge of evolution again worked by indirection to use this product of evolution, the first technology creating species to survive, to create the next stage: technology, a continuation of biological evolution by other means.

The Law of Accelerating Returns

The first stages of technologies, like stone tools, fire, and the wheel took tens of thousands of years, but then we had more powerful tools to create the next stage. A thousand years ago, a paradigm shift like the printing press took only a century or so to be adopted, and this evolution has accelerated ever since. Fifty years ago, the first computers were designed with pencil on paper, with screwdrivers and wire. Today we have computers to design computers. Computer designers will design some high-level parameters, and twelve levels of intermediate design are computed automatically. The process of designing a computer now goes much more quickly.

Evolutionary processes accelerate, and the returns from an evolutionary process grow in power. I've called this theory "The Law of Accelerating Returns." The returns, including economic returns, accelerate. Stemming from my interest in being an inventor, I've been developing mathematical models of this because I quickly realized that an invention has to make sense when the technology is finished, not when it was started, since the world is generally a different place three or four years later.

One exponential pattern that people are familiar with is Moore's Law, which is really just one specific paradigm of shrinking transistors on integrated circuits. It's remarkable how long it's lasted, but it wasn't the first, but the fifth paradigm to provide exponential growth to computing. Earlier, we had electro-mechanical calculators, using relays and vacuum tubes. Engineers were shrinking the vacuum tubes, making them smaller and smaller, until finally that paradigm ran out of steam because they couldn't keep the vacuum any more. Transistors were already in use in radios and other small, niche applications, but when the mainstream technology of computing finally ran out of steam, it switched to this other technology that was already waiting in the

wings to provide ongoing exponential growth. It was a paradigm shift. Later, there was a shift to integrated circuits, and at some point, integrated circuits will run out of steam.

Ten or 15 years from now we'll go to the third dimension. Of course, research on three-dimensional computing is well under way, because as the end of one paradigm becomes clear, this perception increases the pressure for the research to create the next. We've seen tremendous acceleration of molecular computing in the last several years.

When my book, *The Age of Spiritual Machines*, came out about four years ago, the idea that three-dimensional molecular computing could be feasible was quite controversial, and a lot of computer scientists didn't believe it was. Today, there is a universal belief that it's feasible, and that it will arrive in plenty of time before Moore's Law runs out. We live in a three-dimensional world, so we might as well use the third dimension. That will be the sixth paradigm.

Moore's Law is one paradigm among many that have provided exponential growth in computation, but computation is not the only technology that has grown exponentially. We see something similar in any technology, particularly in ones that have any relationship to information.

The genome project, for example, was not a mainstream project when it was announced. People thought it was ludicrous that you could scan the genome in 15 years, because at the rate at which you could scan it when the project began, it could take thousands of years. But the scanning has doubled in speed every year, and actually most of the work was done in the last year of the project.

Magnetic data storage is not covered under Moore's Law, since it involves packing information on a magnetic substrate, which is a completely different set of applied physics, but magnetic data storage has very regularly doubled every year. In fact there's a second level of acceleration. It took us three years to double the price-performance of computing at the beginning of the century, and two years in the middle of the century, but we're now doubling it in less than one year.

This is another feedback loop that has to do with past technologies, because as we improve the price performance, we put more resources into that technology. If you plot computers, as I've done, on a logarithmic scale, where a straight line would mean exponential growth, you see another exponential. There's actually a double rate of exponential growth.

Another very important phenomenon is the rate of paradigm shift. This is harder to measure, but even though people can argue about some of the details and assumptions in these charts you still get these same very powerful trends. The paradigm shift rate itself is accelerating, and roughly doubling every decade. When people claim that we won't see a particular development for a hundred years, or that something is going to take centuries to do accomplish, they're ignoring the inherent acceleration of technical progress.

Bill Joy and I were at Harvard some months ago and one Nobel Prize-winning biologist said that we won't see self-replicating nanotechnology entities for a hundred years. That's actually a good intuition, because that's my estimation—at today's rate of progress—of how long it will take to

achieve that technical milestone. However, since we're doubling the rate of progress every decade, it'll only take 25 calendar years to get there—this, by the way, is a mainstream opinion in the nanotechnology field.

The last century is not a good guide to the next, in the sense that it made only about 20 years of progress at today's rate of progress, because we were speeding up to this point. At today's rate of progress, we'll make the same amount of progress as what occurred in the 20th century in 14 years, and then again in 7 years. The 21st century will see, because of the explosive power of exponential growth, something like 20,000 years of progress at today's rate of progress—a thousand times greater than the 20th century, which was no slouch for radical change.

I've been developing these models for a few decades, and made a lot of predictions about intelligent machines in the 1980s that people can check out. They weren't perfect, but were a pretty good road map. I've been refining these models. I don't pretend that anybody can see the future perfectly, but the power of the exponential aspect of the evolution of these technologies, or of evolution itself, is undeniable. And that creates a very different perspective about the future.

Let's take computation. Communication is important and shrinkage is important. Right now, we're shrinking technology, apparently both mechanical and electronic, at a rate of 5.6 per linear dimension per decade. That number is also moving slowly, in a double exponential sense, but we'll get to nanotechnology at that rate in the 2020s. There are some early-adopter examples of nanotechnology today, but the real mainstream, where the cutting edge of the operating principles are in the multi-nanometer range, will be in the 2020s. If you put these together you get some interesting observations.

Right now we have 10^{26} calculations per second in human civilization in our biological brains. We could argue about this figure, but it's basically, for all practical purposes, fixed. I don't know how much intelligence it adds if you include animals, but maybe you then get a little bit higher than 10^{26}. Non-biological computation is growing at a double exponential rate, and right now is millions of times less than the biological computation in human beings. Biological intelligence is fixed, because it's an old, mature paradigm, but the new paradigm of non-biological computation and intelligence is growing exponentially. The crossover will be in the 2020s and after that, at least from a hardware perspective, non-biological computation will dominate at least quantitatively.

This brings up the question of software. Lots of people say that even though things are growing exponentially in terms of hardware, we've made no progress in software. But we are making progress in software, even if the doubling factor is much slower.

Reverse-engineering the brain

The real scenario that I want to address is the reverse-engineering of the human brain. Our knowledge of the human brain and the tools we have to observe and understand it are themselves growing exponentially. Brain scanning and mathematical models of neurons and neural structures are growing exponentially, and there's very interesting work going on.

There is Lloyd Watts, for example, who with his colleagues has collected models of specific types of neurons and wiring information about how the internal connections are wired in different regions of the brain. He has put together a detailed model of about 15 regions that deal with auditory processing, and has applied psychoacoustic tests of the model, comparing it to human auditory perception.

The model is at least reasonably accurate, and this technology is now being used as a front end for speech recognition software. Still, we're at the very early stages of understanding the human cognitive system. It's comparable to the genome project in its early stages in that we also knew very little about the genome in its early stages. We now have most of the data, but we still don't have the reverse engineering to understand how it works.

It would be a mistake to say that the brain only has a few simple ideas and that once we can understand them we can build a very simple machine. But although there is a lot of complexity to the brain, it's also not vast complexity. It is described by a genome that doesn't have that much information in it. There are about 800 million bytes in the uncompressed genome. We need to consider redundancies in the DNA, as some sequences are repeated hundreds of thousands of times. By applying routine data compression, you can compress this information at a ratio of about 30 to 1, giving you about 23 million bytes—which is smaller than Microsoft Word—to describe the initial conditions of the brain.

But the brain has a lot more information than that. You can argue about the exact number, but I come up with thousands of trillions of bytes of information to characterize what's in a brain, which is millions of times greater than what is in the genome. How can that be?

Marvin talked about how the methods from computer science are important for understanding how the brain works. We know from computer science that we can very easily create programs of considerable complexity from a small starting condition. You can, with a very small program, create a genetic algorithm that simulates some simple evolutionary process and create something of far greater complexity than itself. You can use a random function within the program, which ultimately creates not just randomness, but is creating some meaningful information after the initial random conditions are evolved using a self organizing method, resulting in information that's far greater than the initial conditions.

That is in large measure how the genome creates the brain. We know that it specifies certain constraints for how a particular region is wired, but within those constraints and methods, there's a great deal of stochastic or random wiring, followed by some kind of process whereby the brain learns and self-organizes to make sense of its environment. At this point, what began as random becomes meaningful, and the program has multiplied the size of its information.

The point of all of this is that, since it's a level of complexity we can manage, we will be able to reverse-engineer the human brain. We've shown that we can model neurons, clusters of neurons, and even whole brain regions. We are well down that path. It's rather conservative to say that within 25 years we'll have all of the necessary scanning information and neuron models and will be able to put together a model of the principles of operation of how the human brain works.

Then, of course, we'll have an entity that has some human like qualities. We'll have to educate and train it, but of course we can speed up that process, since we'll have access to everything that's out in the Web, which will contain all accessible human knowledge.

One of the nice things about computer technology is that once you master a process it can operate much faster. So we will learn the secrets of human intelligence, partly from reverse-engineering of the human brain. This will be one source of knowledge for creating the software of intelligence.

We can then combine some advantages of human intelligence with advantages that we see clearly in non-biological intelligence. We spent years training our speech recognition system, which gives us a combination of rules. It mixes expert-system approaches with some self-organizing techniques like neural nets, Markov models and other self-organizing algorithms. We automate the training process by recording thousands of hours of speech and annotating it, and it automatically readjusts all its Markov-model levels and other parameters when it makes mistakes. Finally, after years of this process, it does a pretty good job of recognizing speech. Now, if you want your computer to do the same thing, you don't have to go through those years of training like we do with every child, you can actually load the evolved pattern of this one research computer, which is called loading the software.

Machines can share their knowledge. Machines can do things quickly. Machines have a type of memory that's more accurate than our frail human memories. Nobody at this table can remember billions of things perfectly accurately and look them up quickly. The combination of the software of biological human intelligence with the benefits of non-biological intelligence will be very formidable. Ultimately, this growing non-biological intelligence will have the benefits of human levels of intelligence in terms of its software and our exponentially growing knowledge base.

Superintelligence in the universe

In the future, maybe only one part of intelligence in a trillion will be biological, but it will be infused with human levels of intelligence, which will be able to amplify itself because of the powers of non-biological intelligence to share its knowledge. How does it grow? Does it grow in or does it grow out? Growing in means using finer and finer granularities of matter and energy to do computation, while growing out means using more of the stuff in the universe.

Presently, we see some of both. We see mostly the "in," since Moore's Law inherently means that we're shrinking the size of transistors and integrated circuits, making them finer and finer. To some extent we're also expanding out in that even though the chips are more and more powerful, we make more chips every year, and deploy more economic and material resources towards this non biological intelligence.

Ultimately, we'll get to nanotechnology-based computation, which is at the molecular level, infused with the software of human intelligence and the expanding knowledge base of human civilization. It'll continue to expand both inwards and outwards. It goes in waves as the expansion inwards reaches certain points of resistance. The paradigm shifts will be pretty smooth as we go from the second to the third dimension via molecular computing. At that point it'll be

feasible to take the next step into femto engineering—on the scale of trillionths of a meter—and pico engineering—on the scale of thousands of trillionths of a meter—going into the finer structures of matter and manipulating some of the really fine forces, such as strings and quarks.

That's going to be a barrier, however, so the ongoing expansion of our intelligence is going to be propelled outward. Nonetheless, it will go both in and out. Ultimately, if you do the math, we will completely saturate our corner of the universe, the earth and solar system, sometime in the 22nd century. We'll then want ever-greater horizons, as is the nature of intelligence and evolution, and will then expand to the rest of the universe.

How quickly will it expand? One premise is that it will expand at the speed of light, because that's the fastest speed at which information can travel. There are also tantalizing experiments on quantum disentanglement that show some effect at rates faster than the speed of light, even much faster, perhaps theoretically instantaneously. Interestingly enough, though, this is not the transmission of information, but the transmission of profound quantum randomness, which doesn't accomplish our purpose of communicating intelligence. You need to transmit information, not randomness. So far nobody has actually shown true transmission of information at faster than the speed of light, at least not in a way that has convinced mainstream scientific opinion.

If, in fact, that is a fundamental barrier, and if things that are far away really are far away, which is to say there are no shortcuts through wormholes through the universe, then the spread of our intelligence will be slow, governed by the speed of light. This process will be initiated within 200 years. If you do the math, we will be at near saturation of the available matter and energy in and around our solar system, based on current understandings of the limitations of computation, within that time period.

However, it's my conjecture that by going through these other dimensions that Alan and Paul talked about, there may be shortcuts. It may be very hard to do, but we're talking about supremely intelligent technologies and beings. If there are ways to get to parts of the universe through shortcuts such as wormholes, they'll find, deploy, and master them, and get to other parts of the universe faster. Then perhaps we can reach the whole universe, say 10^{80} protons, photons, and other particles that Seth Lloyd estimates represents on the order of 10^{90} bits, without being limited by the apparent speed of light.

If the speed of light is not a limit, and I do have to emphasize that this particular point is a conjecture at this time, then within 300 years, we would saturate the whole universe with our intelligence, and the whole universe would become supremely intelligent and be able to manipulate everything according to its will. We're currently multiplying computational capacity by a factor of at least 10^3 every decade. This is conservative, as this rate of exponential growth is itself growing exponentially. Thus it is conservative to project that within 30 decades (300 years), we would multiply current computational capacities by a factor of 10^{90}, and thus exceed Seth Lloyd's estimate of 10^{90} bits in the Universe.

We can speculate about identity—will this be multiple people or beings, or one being, or will we all be merged?—but nonetheless, we'll be very intelligent and we'll be able to decide whether we

want to continue expanding. Information is very sacred, which is why death is a tragedy. Whenever a person dies, you lose all that information in a person. The tragedy of losing historical artifacts is that we're losing information. We could realize that losing information is bad, and decide not to do that any more. Intelligence will have a profound effect on the cosmological destiny of the universe at that point.

Why SETI will fail

I'll end with a comment about the SETI project. Regardless of this ultimate resolution of this issue of the speed of light—and it is my speculation (and that of others as well) that there are ways to circumvent it—if there are ways, they'll be found, because intelligence is intelligent enough to master any mechanism that is discovered. Regardless of that, I think the SETI project will fail—it's actually a very important failure, because sometimes a negative finding is just as profound as a positive finding—for the following reason: we've looked at a lot of the sky with at least some level of power, and we don't see anybody out there.

The SETI assumption is that even though it's very unlikely that there is another intelligent civilization like we have here on Earth, there are billions of trillions of planets. So even if the probability is one in a million, or one in a billion, there are still going to be millions, or billions, of life-bearing and ultimately intelligence-bearing planets out there.

If that's true, they're going to be distributed fairly evenly across cosmological time, so some will be ahead of us, and some will be behind us. Those that are ahead of us are not going to be ahead of us by only a few years. They're going to be ahead of us by billions of years. But because of the exponential nature of evolution, once we get a civilization that gets to our point, or even to the point of Babbage, who was messing around with mechanical linkages in a crude 19th century technology, it's only a matter of a few centuries before they get to a full realization of nanotechnology, if not femto and pico-engineering, and totally infuse their area of the cosmos with their intelligence. It only takes a few hundred years!

So if there are millions of civilizations that are millions or billions of years ahead of us, there would have to be millions that have passed this threshold and are doing what I've just said, and have really infused their area of the cosmos. Yet we don't see them, nor do we have the slightest indication of their existence, a challenge known as the Fermi paradox. Someone could say that this "silence of the cosmos" is because the speed of light is a limit, therefore we don't see them, because even though they're fantastically intelligent, they're outside of our light sphere. Of course, if that's true, SETI won't find them, because they're outside of our light sphere.

But let's say they're inside our light sphere, or that light isn't a limitation, for the reasons I've mentioned. Then perhaps they decided, in their great wisdom, to remain invisible to us. You can imagine that there's one civilization out there that made that decision, but are we to believe that this is the case for every one of the millions, or billions, of civilizations that SETI says should be out there?

That's unlikely, but even if it's true, SETI still won't find them, because if a civilization like that has made that decision, it is so intelligent they'll be able to carry that out, and remain hidden

from us. Maybe they're waiting for us to evolve to that point and then they'll reveal themselves to us. Still, if you analyze this more carefully, it's very unlikely in fact that they're out there.

You might ask, isn't it incredibly unlikely that this planet, which is in a very random place in the universe and one of trillions of planets and solar systems, is ahead of the rest of the universe in the evolution of intelligence? Of course the whole existence of our universe, with the laws of physics so sublimely precise to allow this type of evolution to occur is also very unlikely, but by the anthropic principle, we're here, and by an analogous anthropic principle we are here in the lead. After all, if this were not the case, we wouldn't be having this conversation. So by a similar anthropic principle we're able to appreciate this argument.

I'll end on that note.

Deep Fritz Draws: Are Humans Getting Smarter, or Are Computers Getting Stupider?

Ray Kurzweil

http://www.kurzweilai.net/articles/art0527.html

The Deep Fritz computer chess software only achieved a draw in its recent chess tournament with Vladimir Kramnik because it has available only about 1.3% as much brute force computation as the earlier Deep Blue's specialized hardware. Despite that, it plays chess at about the same level because of its superior pattern recognition-based pruning algorithm. In six years, a program like Deep Fritz will again achieve Deep Blue's ability to analyze 200 million board positions per second. Deep Fritz-like chess programs running on ordinary personal computers will routinely defeat all humans later in this decade.

Published on KurzweilAI.net October 19, 2002

In The Age of Intelligent Machines (MIT Press, 1990), which I wrote in 1986-1989, I predicted that a computer would defeat the human world chess champion by the end of the 1990s. I also noted that computers were gaining about 45 points per year in their chess ratings whereas the best human playing was essentially fixed, which projected the cross-over point at 1998. Indeed, Deep Blue did defeat Gary Kasparov in a highly publicized tournament in 1997.

Now with yesterday's final game, we have the current reigning computer program, Deep Fritz, able only to achieve a 4-4 tournament tie with world chess champion Vladimir Kramnik. It has been five years since Deep Blue's victory, so what are we to make of this situation? Should we conclude that:

- Humans are getting smarter, or at least better at chess?
- Computers are getting worse at chess?

And if we were to accept the latter, should we conclude that:

- The much-publicized improvement in computer speed over the past five years was not all it was cracked up to be? Or,
- Computer software is getting worse, at least in chess?

The specialized-hardware advantage

None of the above conclusions are warranted. To gain some insight into these questions, we need to examine a few essentials beneath the surface of the headlines. When I wrote my predictions of computer chess in the late 1980s, Carnegie Mellon University was embarked on a program to

develop specialized chips for conducting the "minimax" algorithm (the standard game-playing method that relies on building trees of move-countermove sequences, and then evaluating the "terminal leaf" positions of each branch of the tree) specifically for chess moves.

Based on this specialized hardware, their 1988 chess machine HiTech was able to analyze 175,000 board positions per second and achieved a chess rating of 2,359, only about 440 points below the human world champion.

A year later in 1989, CMU's "Deep Thought" increased this capacity to 1 million board positions per second and achieved a chess rating of 2,400. IBM eventually took over the project and renamed it "Deep Blue," but kept the basic CMU architecture. The version of Deep Blue that defeated Gary Kasparov in 1997 had 256 special purpose chess processors working in parallel, which analyzed 200 million board positions per second.

An important point to note here was the use of specialized hardware to accelerate the specific calculations needed to generate the minimax algorithm for chess moves. It is well known to computer systems designers that specialized hardware generally can implement a specific algorithm at least 100 times faster than programming the same algorithm as conventional software on a general-purpose computer. ASICs (Application-Specific Integrated Circuits) require significant development efforts and costs, but for critical calculations that are needed on a repetitive basis (for example, decoding MP3 files or rendering graphics primitives for video games), this expenditure can be well worth the investment.

Deep Blue vs. Deep Fritz

Prior to the time when computers could defeat the best human players, there was a great deal of focus on this milestone, so there was support for investing in special-purpose chess circuits. Despite some level of controversy regarding the rules and procedures of the Deep Blue-Kasparov match, the level of interest in computer chess waned considerably after 1997. After all, the goal had been achieved, and there was little point in beating a dead horse. IBM cancelled work on the project, and there has been no work on specialized chess chips since that time.

Computer hardware has nonetheless continued its exponential increase in speed. Personal computer speeds have doubled every year since 1997. Thus the general-purpose Pentium processors used by Deep Fritz are about 32 times faster than personal computer processors back in 1997. Deep Fritz uses a network of only eight personal computers, so the hardware is equivalent to 256 1997-class personal computers.

Compare that to Deep Blue, which used 256 specialized chess processors, each of which were about 100 times faster than 1997 personal computers (of course, only for computing chess minimax). So Deep Blue was 25,600 times faster than a 1997-class personal computer for computing chess moves, and 100 times faster than Deep Fritz. This analysis is confirmed by the reported speeds of the two systems: Deep Blue can analyze 200 million board positions per second compared to only about 2.5 million for Deep Fritz.

Thus the primary problem with Deep Fritz is that it is much slower than Deep Blue. However, the reason for this is the use of specialized hardware in Deep Blue, and the lack of it in Deep Fritz. This reflects the relatively low priority we've given to chess machines since 1997. The focus of research in the various domains spun out of artificial intelligence has been placed instead on problems of greater consequence, such as guiding airplanes, missiles, and factory robots, understanding natural language, diagnosing electrocardiograms and blood cell images, detecting credit card fraud, and a myriad of other successful "narrow" AI applications.

Significant software gains

So what can we say about the software in Deep Fritz? Although chess machines are usually referred to as examples of brute-force calculation, there is one important aspect of these systems that does require qualitative judgment. The combinatorial explosion of possible move-countermove sequences is rather formidable.

In The Age of Intelligent Machines, I estimated that it would take about 40 billion years to make one move if we failed to prune the move-countermove tree and attempted to make a "perfect" move in a typical game (assuming about 30 moves in a typical game and about eight possible moves per play, we have 8^{30} possible move sequences; analyzing one billion of these per second would take 10^{18} seconds or 40 billion years). I noted that this would not be regulation play, so a practical system needs to continually prune away unpromising lines of play. This requires insight and is essentially a pattern-recognition judgment.

Humans, even world class chess masters, perform the minimax algorithm extremely slowly, generally performing less than one move-countermove analysis per second. So how is it that a chess master can compete at all with computer systems that do this millions of times faster? The answer is that we possess formidable powers of pattern recognition. Pattern recognition incidentally is my principal area of technical interest and expertise, and is, in my view, the primary basis of human intelligence. Thus we perform the task of pruning the tree with great insight.

After the Deep Blue-Kasparov match, I suggested to Murray Campbell, head of IBM's Deep Blue team, that they replace the somewhat ad hoc set of rules they used for this pruning judgment task, and replace it with a well-designed neural net. All of the master games of this century are available on line, so it would be possible to train these neural nets on a considerable corpus of expert decisions.

This approach would combine the natural advantage of machines in terms of computational speed with at least a modest step towards more sophisticated pattern recognition. Campbell liked the idea and we were getting set to convene an advisory group to flesh out the idea when IBM cancelled the project.

It is precisely in this area of applying pattern recognition to the crucial pruning decision that Deep Fritz has improved considerably over Deep Blue. Despite Deep Fritz having available only about 1.3% as much brute force computation, it plays chess at about the same level because of its superior pattern-recognition-based pruning algorithm.

So chess software has made significant gains. Deep Fritz has only slightly more computation available than CMU's Deep Thought, yet is rated almost 400 points higher.

Are human chess players doomed?

Another prediction I made in The Age of Intelligent Machines was that once computers did perform as well or better as humans in chess, we would either think more of computer intelligence, or less of human intelligence, or less of chess, and that if history is a guide, we would think less of chess. Indeed, that is what happened. Right after Deep Blue's victory, we heard a lot about how chess is really just a simple game of calculating combinations, and that the computer victory just demonstrated that it was a better calculator.

The reality is slightly more complex. The ability of humans to perform well in chess is clearly not due to our calculating prowess, which we are in fact rather poor at. We use instead a quintessentially human form of judgment. For this type of qualitative judgment, Deep Fritz represents genuine progress over earlier systems.

Incidentally, humans have made no progress in the last five years, with the top human scores remaining just below 2,800. Kasparov is rated at 2,795 and Kramnik at 2,794.

Where we go from here? Now that computer chess is relying on software running on ordinary personal computers, they will continue to benefit from the ongoing acceleration of computer power. In six years, a program like Deep Fritz will again achieve the ability to analyze 200 million board positions per second that was provided by Deep Blue's specialized hardware. With the opportunity to harvest computation on the Internet, we will be able to achieve this potential several years sooner (Internet harvesting of computers will require more ubiquitous broadband communication, but that's coming too).

With these inevitable speed increases, as well as ongoing improvements in pattern recognition, computer chess ratings will continue to edge higher. Deep Fritz-like chess programs running on ordinary personal computers will routinely defeat all humans later in this decade. Then we'll really lose interest in chess.

A Wager on the Turing Test: The Rules

Ray Kurzweil, Mitch Kapor

http://www.kurzweilai.net/articles/art0373.html

An explanation of rules behind the Turing Test, used to determine the winner of a long bet between Ray Kurzweil and Mitch Kapor over whether artificial intelligence will be achieved by 2029.

Published on KurzweilAI.net April 9, 2002. Click here *to see why* ***Ray Kurzweil*** *thinks he will win. Click* here *to read why* ***Mitch Kapor*** *thinks he'll win. Finally, see Ray's* response*.*

Background on the "Long Now Turing Test Wager." Ray Kurzweil maintains that a computer (i.e., a machine intelligence) will pass the Turing test by 2029. Mitchell Kapor believes this will not happen.

This wager is intended to be the inaugural long term bet to be administered by the Long Now Foundation. The proceeds of the wager are to be donated to a charitable organization designated by the winner.

This document provides a brief description of the Turing Test and a set of high level rules for administering the wager. These rules contemplate setting up a "Turing Test Committee" which will create the detailed rules and procedures to implement the resolution of the wager. A primary objective of the Turing Test Committee will be to set up rules and procedures that avoid and deter cheating.

Brief Description of the Turing test. In a 1950 paper ("Computing Machinery and Intelligence," *Mind 59* (1950): 433-460, reprinted in E. Feigenbaum and J. Feldman, eds., *Computers and Thought*, New York: McGraw-Hill, 1963), Alan Turing describes his concept of the Turing Test, in which one or more human judges interview computers and human foils using terminals (so that the judges won't be prejudiced against the computers for lacking a human appearance). The nature of the dialog between the human judges and the candidates (i.e., the computers and the human foils) is similar to an online chat using instant messaging. The computers as well as the human foils try to convince the human judges of their humanness. If the human judges are unable to reliably unmask the computers (as imposter humans) then the computer is considered to have demonstrated human-level intelligence[1].

Turing was very specifically nonspecific about many aspects of how to administer the test. He did not specify many key details, such as the duration of the interrogation and the sophistication of the human judge and foils. The purpose of the rules described below is to provide a set of procedures for administering the test some decades hence.

The Procedure for the Turing Test Wager: The Turing Test General Rules

These Turing Test General Rules may be modified by agreement of Ray Kurzweil and Mitchell Kapor, or, if either Ray Kurzweil and / or Mitchell Kapor is not available, then by the Turing Test Committee (described below). However, any such change to these Turing Test General Rules shall only be made if (i) these rules are determined to have an inconsistency, or (ii) these rules are determined to be inconsistent with Alan Turing's intent of determining human-level intelligence in a machine, or (iii) these rules are determined to be unfair, or (iv) these rules are determined to be infeasible to implement.

I. Definitions.

A Human is a biological human person as that term is understood in the year 2001 whose intelligence has not been enhanced through the use of machine (i.e., nonbiological) intelligence, whether used externally (e.g., the use of an external computer) or internally (e.g., neural implants). A Human may not be genetically enhanced (through the use of genetic engineering) beyond the level of human beings in the year 2001.

A Computer is any form of nonbiological intelligence (hardware and software) and may include any form of technology, but may not include a biological Human (enhanced or otherwise) nor biological neurons (however, nonbiological emulations of biological neurons are allowed).

The Turing Test Committee will consist of three Humans, to be selected as described below.

The Turing Test Judges will be three Humans selected by the Turing Test Committee.

The Turing Test Human Foils will be three Humans selected by the Turing Test Committee.

The Turing Test Participants will be the three Turing Test Human Foils and one Computer.

II. The Procedure

The Turing Test Committee will be appointed as follows.

- One member will be Ray Kurzweil or his designee, or, if not available, a person appointed by the Long Now Foundation. In the event that the Long Now Foundation appoints this person, it shall use its best efforts to appoint a Human person that best represents the views of Ray Kurzweil (as expressed in the attached essay "Why I Think I Will Win The Long Now Turing Test Wager.")
- A second member will be Mitchell Kapor or his designee, or, if not available, a person appointed by the Long Now Foundation. In the event that the Long Now Foundation appoints this person, it shall use its best efforts to appoint a Human person that best represents the views of Mitchell Kapor (as expressed in the attached essay "Why I Think I Will Win The Long Now Turing Test Wager.")
- A third member will be appointed by the above two members, or if the above two members are unable to agree, then by the Long Now Foundation, who in its judgment, is qualified to represent a "middle ground" position.

Ray Kurzweil, or his designee, or another member of the Turing Test Committee, or the Long Now Foundation may, from time to time call for a Turing Test Session to be conducted and will select or provide one Computer for this purpose. For those Turing Test Sessions called for by Ray Kurzweil or his designee or another member of the Turing Test committee (other than the final one in 2029), the person calling for the Turing Test Session to be conducted must provide (or raise) the funds necessary for the Turing Test Session to be conducted. In any event, the Long Now Foundation is not obligated to conduct more than two such Turing Test Sessions prior to the final one (in 2029) if it determines that conducting such additional Turing Test Sessions would be an excessive administrative burden.

The Turing Test Committee will provide the detailed rules and procedures to implement each such Turing Test Session using its best efforts to reflect the rules and procedures described in this document. The primary goal of the Turing Test Committee will be to devise rules and procedures which avoid and deter cheating to the maximum extent possible. These detailed rules and procedures will include (i) specifications of the equipment to be used, (ii) detailed procedures to be followed, (iii) specific instructions to be given to all participants including the Turing Test Judges, the Turing Test Human Foils and the Computer, (iv) verification procedures to assure the integrity of the proceedings, and (v) any other details needed to implement the Turing Test Session. Beyond the Turing Test General Rules described in this document, the Turing Test Committee will be guided to the best of its ability by the original description of the Turing Test by Alan Turing in his 1950 paper. The Turing Test Committee will also determine procedures to resolve any deadlocks that may occur in its own deliberations.

Each Turing Test Session will consist of at least three Turing Test Trials.

For each such Turing Test Trial, a set of Turing Test Interviews will take place, followed by voting by the Turing Test Judges as described below.

Using its best judgment, the Turing Test Committee will appoint three Humans to be the Turing Test Judges.

Using its best judgment, the Turing Test Committee will appoint three Humans to be the Turing Test Human Foils. The Turing Test Human Foils should not be known (either personally or by reputation) to the Turing Test Judges.

During the Turing Test Interviews (for each Turing Test Trial), each of the three Turing Test Judges will conduct online interviews of each of the four Turing Test Candidates (i.e., the Computer and the three Turing Test Human Foils) for two hours each for a total of eight hours of interviews conducted by each of the three Turing Test Judges (for a total of 24 hours of interviews).

The Turing Test Interviews will consist of online text messages sent back and forth as in a online "instant messaging" chat, as that concept is understood in the year 2001.

The Human Foils are instructed to try to respond in as human a way as possible during the Turing Test Interviews.

The Computer is also intended to respond in as human a way as possible during the Turing Test Interviews.

Neither the Turing Test Human Foils nor the Computer are required to tell the truth about their histories or other matters. All of the candidates are allowed to respond with fictional histories.

At the end of the interviews, each of the three Turing Test Judges will indicate his or her verdict with regard to each of the four Turing Test Candidates indicating whether or not said candidate is human or machine. The Computer will be deemed to have passed the "Turing Test Human Determination Test" if the Computer has fooled two or more of the three Human Judges into thinking that it is a human.

In addition, each of the three Turing Test Judges will rank the four Candidates with a rank from 1 (least human) to 4 (most human). The computer will be deemed to have passed the "Turing Test Rank Order Test" if the median rank of the Computer is equal to or greater than the median rank of two or more of the three Turing Test Human Foils.

The Computer will be deemed to have passed the Turing Test if it passes both the Turing Test Human Determination Test and the Turing Test Rank Order Test.

If a Computer passes the Turing Test, as described above, prior to the end of the year 2029, then Ray Kurzweil wins the wager. Otherwise Mitchell Kapor wins the wager.

[1] Turing's initial description of his test was as a parlor game in which judges try to determine the gender of male and female human contestants. He then suggests the applicability of this type of game to its present purpose of determining when the level of intelligence of a machine is indistinguishable from that of a human.

A Wager on the Turing Test: Why I Think I Will Win

Ray Kurzweil

http://www.kurzweilai.net/articles/art0374.html

Will Ray Kurzweil's predictions come true? He's putting his money where his mouth is. Here's why he thinks he will win a bet on the future of artificial intelligence. The wager: an AI that passes the Turing Test by 2029.

Published April 9, 2002on KurzweilAI.net. Click here *to read an explanation of the bet and its background, with rules and definitions. Click* here *to read* ***Mitch Kapor****'s response. Also see* ***Ray Kurzweil****'s* final word *on why he will win.*

The Significance of the Turing Test. The implicit, and in my view brilliant, insight in Turing's eponymous test is the ability of written human language to represent human-level thinking. The basis of the Turing test is that if the human Turing test judge is competent, then an entity requires human-level intelligence in order to pass the test. The human judge is free to probe each candidate with regard to their understanding of basic human knowledge, current events, aspects of the candidate's personal history and experiences, as well as their subjective experiences, all expressed through written language. As humans jump from one concept and one domain to the next, it is possible to quickly touch upon all human knowledge, on all aspects of human, well, humanness.

To the extent that the "AI" chooses to reveal its "history" during the interview with the Turing Test judge (note that none of the contestants are required to reveal their histories), the AI will need to use a fictional human history because "it" will not be in a position to be honest about its origins as a machine intelligence and pass the test. (By the way, I put the word "it" in quotes because it is my view that once an AI does indeed pass the Turing Test, we may very well consider "it" to be a "he" or a "she.") This makes the task of the machines somewhat more difficult than that of the human foils because the humans can use their own history. As fiction writers will attest, presenting a totally convincing human history that is credible and tracks coherently is a challenging task that most humans are unable to accomplish successfully. However, some humans are capable of doing this, and it will be a necessary task for a machine to pass the Turing test.

There are many contemporary examples of computers passing "narrow" forms of the Turing test, that is, demonstrating human-level intelligence in specific domains. For example, Gary Kasparov, clearly a qualified judge of human chess intelligence, declared that he found Deep Blue's playing skill to be indistinguishable from that of a human chess master during the famous tournament in which he was defeated by Deep Blue. Computers are now displaying human-level intelligence in a growing array of domains, including medical diagnosis, financial investment decisions, the design of products such as jet engines, and a myriad of other tasks that previously required humans to accomplish. We can say that such "narrow AI" is the threshold that the field of AI has currently achieved. However, the subtle and supple skills required to pass the broad

Turing test as originally described by Turing is far more difficult than any narrow Turing Test. In my view, there is no set of tricks or simpler algorithms (i.e., methods simpler than those underlying human level intelligence) that would enable a machine to pass a properly designed Turing test without actually possessing intelligence at a fully human level.

There has been a great deal of philosophical discussion and speculation concerning the issue of consciousness, and whether or not we should consider a machine that passed the Turing test to be conscious. Clearly, the Turing test is not an explicit test for consciousness. Rather, it is a test of human-level performance. My own view is that inherently there is no objective test for subjective experience (i.e., consciousness) that does not have philosophical assumptions built into it. The reason for this has to do with the difference between the concepts of objective and subjective experience. However, it is also my view that once nonbiological intelligence does achieve a fully human level of intelligence, such that it can pass the Turing test, humans will treat such entities as if they were conscious. After all, they (the machines) will get mad at us if we don't. However, this is a political prediction rather than a philosophical position.

It is also important to note that once a computer does achieve a human level of intelligence, it will necessarily soar past it. Electronic circuits are already at least 10 million times faster than the electrochemical information processing in our interneuronal connections. Machines can share knowledge instantly, whereas we biological humans do not have quick downloading ports on our neurotransmitter concentration levels, interneuronal connection patterns, nor any other biological bases of our memory and skill. Language-capable machines will be able to access vast and accurate knowledge bases, including reading and mastering all the literature and sources of information available to our human-machine civilization. Thus "Turing Test level" machines will be able to combine human level intelligence with the powerful ways in which machines already excel. In addition, machines will continue to grow exponentially in their capacity and knowledge. It will be a formidable combination.

Why I Think I Will Win. In considering the question of when machine (i.e., nonbiological) intelligence will match the subtle and supple powers of human biological intelligence, we need to consider two interrelated but distinct questions: when will machines have the *hardware* capacity to match human information processing, and when will our technology have mastered the methods, i.e., the *software* of human intelligence. Without the latter, we would end up with extremely fast calculators, and would not achieve the endearing qualities that characterize human discernment (nor the deep knowledge and command of language necessary to pass a full Turing test!).

Both the hardware and software sides of this question are deeply influenced by the exponential nature of information-based technologies. The exponential growth that we see manifest in "Moore's Law" is far more pervasive than commonly understood. Our first observation is that the shrinking of transistors on an integrated circuit, which is the principle of Moore's Law, was not the first but the fifth paradigm to provide exponential growth to computing (after electromechanical calculators, relay-based computers, vacuum tube-based computing, and discrete transistors). Each time one approach begins to run out of steam, research efforts intensify to find the next source of renewed exponential growth (e.g., vacuum tubes were made smaller until it was no longer feasible to maintain a vacuum, which led to transistors). Thus the

power and price-performance of technologies, particularly information-based technologies, grow as a cascade of S-curves: exponential growth leading to an asymptote, leading to paradigm shift (i.e., innovation), and another S-curve. Moreover, the underlying theory of the exponential growth of information-based technologies, which I call the law of accelerating returns, as well as a detailed examination of the underlying data, show that there is a second level of exponential growth, i.e., the rate of exponential growth is itself growing exponentially[i].

Second, this phenomenon of ongoing exponential growth through a cascade of S-curves is far broader than computation. We see the same double exponential growth in a wide range of technologies, including communication technologies (wired and wireless), biological technologies (e.g., DNA base-pair sequencing), miniaturization, and of particular importance to the software of intelligence, brain reverse engineering (e.g., brain scanning, neuronal and brain region modeling).

Within the next approximately fifteen years, the current computational paradigm of Moore's Law will come to an end because by that time the key transistor features will only be a few atoms in width. However, there are already at least two dozen projects devoted to the next (i.e., the sixth) paradigm, which is to compute in three-dimensions. Integrated circuits are dense but flat. We live in a three-dimensional world, our brains are organized in three dimensions, and we will soon be computing in three dimensions. The feasibility of three-dimensional computing has already been demonstrated in several landmark projects, including the particularly powerful approach of nanotube-based electronics. However, for those who are (irrationally) skeptical of the potential for three-dimensional computing, it should be pointed out that achieving even a conservatively high estimate of the information processing capacity of the human brain (i.e., one hundred billion neurons times a thousand connections per neuron times 200 digitally controlled analog "transactions" per second, or about 20 million billion operations per second) will be achieved by conventional silicon circuits prior to 2020.

It is correct to point out that achieving the "software" of human intelligence is the more salient, and more difficult, challenge. On multiple levels, we are being guided in this effort by a grand project to reverse engineer (i.e., understand the principles of operation of) the human brain itself. Just as the human genome project accelerated (with the bulk of the genome being sequenced in the last year of the project), the effort to reverse engineer the human brain is also growing exponentially, and is further along than most people realize. We already have highly detailed mathematical models of several dozen of the several hundred types of neurons found in the brain. The resolution, bandwidth, and price-performance of human brain scanning are also growing exponentially. By combining the neuron modeling and interconnection data obtained from scanning, scientists have already reverse engineered two dozen of the several hundred regions of the brain. Implementations of these reverse engineered models using contemporary computation matches the performance of the biological regions that were recreated in significant detail. Already, we are in a early stage of being able to replace small regions of the brain that have been damaged from disease or disability using neural implants (e.g., ventral posterior nucleus, subthalmic nucleus, and ventral lateral thalamus neural implants to counteract Parkinson's Disease and tremors from other neurological disorders, cochlear implants, emerging retinal implants, and others).

If we combine the exponential trends in computation, communications, and miniaturization, it is a conservative expectation that we will within 20 to 25 years be able to send tiny scanners the size of blood cells into the brain through the capillaries to observe interneuronal connection data and even neurotransmitter levels from up close. Even without such capillary-based scanning, the contemporary experience of the brain reverse engineering scientists, (e.g., Lloyd Watts, who has modeled over a dozen regions of the human auditory system), is that the connections in a particular region follow distinct patterns, and that it is not necessary to see every connection in order to understand the massively parallel, digital controlled analog algorithms that characterize information processing in each region. The work of Watts and others has demonstrated another important insight, that once the methods in a brain region are understood and implemented using contemporary technology, the computational requirements for the machine implementation requires on the order of a thousand times less computation than the theoretical potential of the biological neurons being simulated.

A careful analysis of the requisite trends shows that we will understand the principles of operation of the human brain and be in a position to recreate its powers in synthetic substrates well within thirty years. The brain is self-organizing, which means that it is created with relatively little innate knowledge. Most of its complexity comes from its own interaction with a complex world. Thus it will be necessary to provide an artificial intelligence with an education just as we do with a natural intelligence. But here the powers of machine intelligence can be brought to bear. Once we are able to master a process in a machine, it can perform its operations at a much faster speed than biological systems. As I mentioned, contemporary electronics is already more than ten million times faster than the human nervous system's electrochemical information processing. Once an AI masters human basic language skills, it will be in a position to expand its language skills and general knowledge by rapidly reading all human literature and by absorbing the knowledge contained on millions of web sites. Also of great significance will be the ability of machines to share their knowledge instantly.

One challenge to our ability to master the apparent complexity of human intelligence in a machine is whether we are capable of building a system of this complexity without the brittleness that often characterizes very complex engineering systems. This a valid concern, but the answer lies in emulating the ways of nature. The initial design of the human brain is of a complexity that we can already manage. The human brain is characterized by a genome with only 23 million bytes of useful information (that's what left of the 800 million byte genome when you eliminate all of the redundancies, e.g., the sequence called "ALU" which is repeated hundreds of thousands of times). 23 million bytes is smaller than Microsoft WORD. How is it, then, that the human brain with its 100 trillion connections can result from a genome that is so small? The interconnection data alone is a million times greater than the information in the genome.

The answer is that the genome specifies a set of processes, each of which utilizes chaotic methods (i.e., initial randomness, then self-organization) to increase the amount of information represented. It is known, for example, that the wiring of the interconnections follows a plan that includes a great deal of randomness. As the individual person encounters her environment, the connections and the neurotransmitter level pattern self-organize to better represent the world, but the initial design is specified by a program that is not extreme in its complexity.

Thus we will not program human intelligence link by link as in some massive expert system. Nor is it the case that we will simply set up a single genetic (i.e., evolutionary) algorithm and have intelligence at human levels automatically evolve itself. Rather we will set up an intricate hierarchy of self-organizing systems, based largely on the reverse engineering of the human brain, and then provide for its education. However, this learning process can proceed hundreds if not thousands of times faster than the comparable process for humans.

Another challenge is that the human brain must incorporate some other kind of "stuff" that is inherently impossible to recreate in a machine. Penrose imagines that the intricate tubules in human neurons are capable of quantum based processes, although there is no evidence for this. I would point out that even if the tubules do exhibit quantum effects, there is nothing barring us from applying these same quantum effects in our machines. After all, we routinely use quantum methods in our machines today. The transistor, for example, is based on quantum tunneling. The human brain is made of the same small list of proteins that all biological systems are comprised of. We are rapidly recreating the powers of biological substances and systems, including neurological systems, so there is little basis to expect that the brain relies on some nonengineerable essence for its capabilities. In some theories, this special "stuff" is associated with the issue of consciousness, e.g., the idea of a human soul associated with each person. Although one may take this philosophical position, the effect is to separate consciousness from the performance of the human brain. Thus the absence of such a soul may in theory have a bearing on the issue of consciousness, but would not prevent a nonbiological entity from the performance abilities necessary to pass the Turing test.

Another challenge is that an AI must have a human or human-like body in order to display human-like responses. I agree that a body is important to provide a situated means to interact with the world. The requisite technologies to provide simulated or virtual bodies are also rapidly advancing. Indeed, we already have emerging replacements or augmentations for virtually every system in our body. Moreover, humans will be spending a great deal of time in full immersion virtual reality environments incorporating all of the senses by 2029, so a virtual body will do just as well. Fundamentally, emulating our bodies in real or virtual reality is a less complex task than emulating our brains.

Finally, we have the challenge of emotion, the idea that although machines may very well be able to master the more analytical cognitive abilities of humans, they inherently will never be able to master the decidedly illogical and much harder to characterize attributes of human emotion. A slightly broader way of characterizing this challenge is to pose it in terms of "qualia," which refers essentially to the full range of subjective experiences. Keep in mind that the Turing test is assessing convincing reactions to emotions and to qualia. The apparent difficulty of responding appropriately to emotion and other qualia appears to be at least a significant part of Mitchell Kapor's hesitation to accept the idea of a Turing-capable machine. It is my view that understanding and responding appropriately to human emotion is indeed the most complex thing that we do (with other types of qualia being if anything simpler to respond to). It is the cutting edge of human intelligence, and is precisely the heart of the Turing challenge. Although human emotional intelligence is complex, it nonetheless remains a capability of the human brain, with our endocrine system adding only a small measure of additional complexity (and operating at a relatively low bandwidth). All of my observations above pertain to the issue of emotion, because

that is the heart of what we are reverse engineering. Thus, we can say that a side benefit of creating Turing-capable machines will be new levels of insight into ourselves.

[i] All of the points addressed in this statement of "Why I Think I Will Win" (the Long Now Turing Test Wager) are examined in more detail in my essay "The Law of Accelerating Returns" available at http://www.kurzweilai.net/meme/frame.html?main=/articles/art0134.html.

Response to Mitchell Kapor's "Why I Think I Will Win"

Ray Kurzweil

http://www.kurzweilai.net/articles/art0413.html

Ray Kurzweil responds to Mitch Kapor's arguments against the possibility that an AI that will pass a Turing Test in 2029 in this final counterpoint on the bet: an AI will pass the Turing Test by 2029.

Published on KurzweilAI.net April 9, 2002. Click here to read an explanation of the bet and its background, with rules and definitions. Read why Ray thinks he will win here. Click here to see why Mitch Kapor thinks he won't.

Mitchell's essay provides a thorough and concise statement of the classic arguments against the likelihood of Turing-level machines in a several decade timeframe. Mitch ends with a nice compliment comparing me to future machines, and I only wish that it were true. I think of all the books and web sites I'd like to read, and of all the people I'd like to dialog and interact with, and I realize just how limited my current bandwidth and attention span is with my mere hundred trillion connections.

I discussed several of Mitchell's insightful objections in my statement, and augment these observations here:

"We are embodied creatures": True, but machines will have bodies also, in both real and virtual reality.

"Emotion is as or more basic than cognition": Yes, I agree. As I discussed, our ability to perceive and respond appropriately to emotion is the most complex thing that we do. Understanding our emotional intelligence will be the primary target of our reverse engineering efforts. There is no reason that we cannot understand our own emotions and the complex biological system that gives rise to them. We've already demonstrated the feasibility of understanding regions of the brain in great detail.

"We are conscious beings, capable of reflection and self-awareness." I think we have to distinguish the performance aspects of what is commonly called consciousness (i.e., the ability to be reflective and aware of ourselves) versus consciousness as the ultimate ontological reality. Since the Turing test is a test of performance, it is the performance aspects of what is commonly referred to as consciousness that we are concerned with here. And in this regard, our ability to build models of ourselves and our relation to others and the environment is indeed a subtle and complex quality of human thinking. However there is no reason why a nonbiological intelligence would be restricted from similarly building comparable models in its nonbiological brain.

Mitchell cites the limitations of the expert system methodology and I agree with this. A lot of AI criticism is really criticism of this approach. The core strength of human intelligence is not logical analysis of rules, but rather pattern recognition, which requires a completely different

paradigm. This pertains also to Mitchell's objection to the "metaphor" of "brain-as-computer." The future machines that I envision will not be like the computers of today, but will be biologically inspired and will be emulating the massively parallel, self-organizing, holographically organized methods that are used in the human brain. A future AI certainly won't be using expert system techniques. Rather, it will be a complex system of systems, each built with a different methodology, just like, well, the human brain.

I will say that Mitchell is overlooking the hundreds of ways in which "narrow AI" has infiltrated our contemporary systems. Expert systems are not the best example of these, and I cited several categories in my statement.

I agree with Mitchell that the brain does not represent the entirety of our thinking process, but it does represent the bulk of it. In particular, the endocrine system is orders of magnitude simpler and operates at very low bandwidth compared to neural processes (which themselves utilize a form of analog information processing dramatically slower than contemporary electronic systems).

Mitchell expresses skepticism that "it's all about the bits and just the bits." There is something going on in the human brain, and these processes are not hidden from us. I agree that it's actually not exactly bits because what we've already learned is that the brain uses digitally controlled analog methods. We know that analog methods can be emulated by digital methods but there are engineering reasons to prefer analog techniques because they are more efficient by several orders of magnitude. However, the work of Cal Tech Professor Carver Mead and others have shown that we can use this approach in our machines. Again, this is different from today's computers, but will be, I believe, an important future trend.

However, I think Mitchell's primary point here is not to distinguish analog and digital computing methods, but to make reference to some other kind of "stuff" that we inherently can't recreate in a machine. I believe, however, that the scale of the human nervous system (and, yes, the endocrine system, although as I said this adds little additional complexity) is sufficient to explain the complexity and subtlety of our behavior.

I think the most compelling argument that Mitchell offers is his insight that most experience is not book learning. I agree, but point out that one of the primary purposes of nonbiological intelligence is to interact with us humans. So embodied AI's will have plenty of opportunity to learn from direct interaction with their human progenitors, as well as to observe a massive quantity of other full immersion human interaction available over the web.

Now it's true that AI's will have a different history from humans, and that does represent an additional challenge to their passing the Turing test. As I pointed out in my statement, it's harder (even for humans) to successfully defend a fictional history than a real one. So an AI will actually need to surpass native human intelligence in order to pass for a human in a valid Turing test. And that's what I'm betting on.

I can imagine Mitchell saying to himself as he reads this "But does Ray really appreciate the extraordinary depth of human intellect and emotion?" I believe that I do and think that Mitchell

has done an excellent job of articulating this perspective. I would put the question back and ask whether Mitchell really appreciates the extraordinary power and depth of the technology that lies ahead, which will be billions of times more powerful and complex than what we have today?

On that note, I would end by emphasizing the accelerating pace of progress in all of these information-based technologies. The power of these technologies is doubling every year, and the paradigm shift rate is doubling every decade, so the next thirty years will be like 140 years at today's rate of progress. And the past 140 years was comparable to only about 30 years of progress at today's rate of progress because we've been accelerating up to this point. If one really absorbs the implications of what I call the law of accelerating returns, then it becomes apparent that over the next three decades (well, 28 years to be exact when Mitchell and I sit down to compare notes), we will see astonishing levels of technological progress.

Will Machines Become Conscious

“Suppose we scan someone’s brain and reinstate the resulting ‘mind file’ into suitable computing medium,” asks Ray Kurzweil. “Will the entity that emerges from such an operation be conscious?”

How Can We Possibly Tell if it's Conscious?

Ray Kurzweil

http://www.kurzweilai.net/articles/art0461.html

At the Tucson 2002: Toward a Science of Consciousness conference, Ray Kurzweil addressed the question of how to tell if something is conscious. He proposed two thought experiments.

Published on KurzweilAI.net April 18, 2002

Thought Experiment 1: Create Ray 2

- Make a copy of all the salient details of me.
- Ray 2 has all of my memories, so he remembers having been me.
- He recalls all of my memories.
- If you meet him, you would be convinced he is Ray (he passes a "Ray Turing Test").
- You could do this while I was sleeping, so I may not even know about Ray 2.
- If you tell me that we don't need Ray 1 anymore, I may beg to differ.
- I may come to believe in Ray 2's existence, but I would consider him "someone else.".
- Ray 2's continued existence does not represent immortality for me.
- Copying me does not transfer my consciousness because I'm still here... Okay, so far so good.

Thought Experiment 2: Gradual Replacement of Ray

- Replace a tiny portion of my brain with its neuromorphic equivalent.
- Okay, I'm still here... the operation was successful (eventually the nanobots will do this without surgery).
- We know people like this already (e.g., people with cochlear implants, Parkinson's implants, other neural implants).
- Do it again... okay I'm still here... and again...
- At the end of the process, I'm still here. There never was an "old Ray" and a "new Ray". I'm the same as I was before. No one ever missed me, including me.
- Gradual replacement of Ray results in Ray, so consciousness and identity appears to have been preserved.
- HOWEVER...

However...

- "Ray at the end of the gradual replacement scenario" is entirely equivalent to Ray 2 in the mental porting scenario.
- "Ray at the end of the gradual replacement scenario" is not Ray but someone else.
- But in the gradual replacement scenario, when did Ray become someone else?

- The gradual replacement scenario is entirely equivalent to what happens naturally:
  – Most of our cells turn over within a month or a few months.
  – Those that persist longer (e.g., neurons) nonetheless replace their particles.
- So are we continually being replaced by someone else?

This is a real issue with regard to Cyronics

- Assuming a "preserved" person is ultimately "reanimated," many of the proposed methods imply that the reanimated person will be "rebuilt" with new materials and even entirely new neuromorphically equivalent systems.
- The reanimated person will, therefore, effectively be "Ray 2" (i.e., someone else).

There is no objective (third party) test for subjectivity (first person experience aka consciousness)

- If Ray 2 happens to be nonbiological:
  – He would have all of the same abilities to understand his own situation, the same feedback loops.
  – The activity in his nonbiological brain would be comparable to Ray 1.
  – He would be completely convincing to Ray's friends.
  – But there is no way to experience his subjective experiences without making philosophical assumptions.
- Machines today are not convincing, but they are still much simpler than human intelligence.
- But this gap will be closed, and future machines will be "convincing" in their emotional intelligence.
- But there remains an inherent objective gap in assessing the subjective experience of another entity.

The "Hard" Issue of Consciousness

- Only becomes a scientific question when one makes certain philosophical assumptions.
- Some people conclude that consciousness is an illusion, that there is no real issue.
- Consciousness is, therefore, the ultimate ontological question, the appropriate province of philosophy and religion.
- So my philosophy is...

... Patternism

- Our ultimate reality is our pattern.
- Knowledge is a pattern as distinguished from mere information.
- Losing knowledge is a profound loss.
  – Losing a person is a profound loss.
- Patterns persist.
  – The pattern of water in a stream
  – Ray

- We are still left with dilemmas because patterns can be copied .

We will ultimately come to accept that nonbiological entities can be (are) conscious

- But this is a political and psychological prediction.
- There is no way to demonstrate this without making philosophical assumptions .

You all seem conscious, but

- Maybe I'm living in a simulation and you're all part of the simulation.
- Or (at the end of the conference), perhaps I only have memories of you, but the experiences never actually took place.
- Or, maybe I am only having a conscious experience of recalling memories of you but neither you nor the memories really exist.
- Or, perhaps...

Ray Kurzweil's The Law of Accelerating Returns essay includes further discussion of these issues about consciousness.

My Question for the Edge: Who am I? What am I?

Ray Kurzweil

http://www.kurzweilai.net/articles/art0376.html

Since we constantly changing, are we just patterns? What if someone copies that pattern? Am I the original and/or the copy? Ray Kurzweil responds to Edge publisher/editor John Brockman's request to futurists to pose "hard-edge" questions that "render visible the deeper meanings of our lives, redefine who and what we are."

Published on KurzweilAI.net January 14, 2002. Originally written January 13, 2002 and published on http://www.edge.org.

Perhaps I am this stuff here, i.e., the ordered and chaotic collection of molecules that comprise my body and brain.

But there's a problem. The specific set of particles that comprise my body and brain are completely different from the atoms and molecules than comprised me only a short while (on the order of weeks) ago. We know that most of our cells are turned over in a matter of weeks. Even those that persist longer (e.g., neurons) nonetheless change their component molecules in a matter of weeks.

So I am a completely different set of stuff than I was a month ago. All that persists is the pattern of organization of that stuff. The pattern changes also, but slowly and in a continuum from my past self. From this perspective I am rather like the pattern that water makes in a stream as it rushes past the rocks in its path. The actual molecules (of water) change every millisecond, but the pattern persists for hours or even years.

So, perhaps we should say I am a pattern of matter and energy that persists in time.

But there is a problem here as well. We will ultimately be able to scan and copy this pattern in a at least sufficient detail to replicate my body and brain to a sufficiently high degree of accuracy such that the copy is indistinguishable from the original (i.e., the copy could pass a "Ray Kurzweil" Turing test). I won't repeat all the arguments for this here, but I describe this scenario in a number of documents including the essay "The Law of Accelerating Returns".

The copy, therefore, will share my pattern. One might counter that we may not get every detail correct. But if that is true, then such an attempt would not constitute a proper copy. As time goes on, our ability to create a neural and body copy will increase in resolution and accuracy at the same exponential pace that pertains to all information-based technologies. We ultimately will be able to capture and recreate my pattern of salient neural and physical details to any desired degree of accuracy.

Although the copy shares my pattern, it would be hard to say that the copy is me because I would (or could) still be here. You could even scan and copy me while I was sleeping. If you come to

me in the morning and say, "Good news, Ray, we've successfully reinstantiated you into a more durable substrate, so we won't be needing your old body and brain anymore," I may beg to differ.

If you do the thought experiment, it's clear that the copy may look and act just like me, but it's nonetheless not me because I may not even know that he was created. Although he would have all my memories and recall having been me, from the point in time of his creation, Ray 2 would have his own unique experiences and his reality would begin to diverge from mine.

Now let's pursue this train of thought a bit further and you will see where the dilemma comes in. If we copy me, and then destroy the original, then that's the end of me because as we concluded above the copy is not me. Since the copy will do a convincing job of impersonating me, no one may know the difference, but it's nonetheless the end of me. However, this scenario is entirely equivalent to one in which I am replaced gradually. In the case of gradual replacement, there is no simultaneous old me and new me, but at the end of the gradual replacement process, you have the equivalent of the new me, and no old me. So gradual replacement also means the end of me.

However, as I pointed out at the beginning of this question, it is the case that I am in fact being continually replaced. And, by the way, it's not so gradual, but a rather rapid process. As we concluded, all that persists is my pattern. But the thought experiment above shows that gradual replacement means the end of me even if my pattern is preserved. So am I constantly being replaced by someone else who just seems a like lot me a few moments earlier?

So, again, who am I? It's the ultimate ontological question. We often refer to this question as the issue of consciousness. I have consciously (no pun intended) phrased the issue entirely in the first person because that is the nature of the issue. It is not a third person question. So my question is not "Who is John Brockman?" although John may ask this question himself.

When people speak of consciousness, they often slip into issues of behavioral and neurological correlates of consciousness (e.g., whether or not an entity can be self-reflective), but these are third person (i.e., objective) issues, and do not represent what David Chalmers calls the "hard question" of consciousness.

The question of whether or not an entity is conscious is only apparent to himself. The difference between neurological correlates of consciousness (e.g., intelligent behavior) and the ontological reality of consciousness is the difference between objective (i.e., third person) and subjective (i.e., first person) reality. For this reason, we are unable to propose an objective consciousness detector that does not have philosophical assumptions built into it.

I do say that we (humans) will come to accept that nonbiological entities are conscious because ultimately they will have all the subtle cues that humans currently possess that we associate with emotional and other subjective experiences. But that's a political and psychological prediction, not an observation that we will be able to scientifically verify. We do assume that other humans are conscious, but this is an assumption, and not something we can objectively demonstrate.

I will acknowledge that John Brockman did seem conscious to me when he interviewed me, but I should not be too quick to accept this impression. Perhaps I am really living in a simulation, and

John was part of the simulation. Or, perhaps it's only my memories that exist, and the actual experience never took place. Or maybe I am only now experiencing the sensation of recalling apparent memories of having met John, but neither the experience nor the memories really exist. Well, you see the problem.

Read other questions and answers at The Edge's World Question Center.

Live Forever—Uploading the Human Brain… Closer Than You Think

Ray Kurzweil

http://www.kurzweilai.net/articles/art0157.html?m=4

Ray Kurzweil ponders the issues of identity and consciousness in an age when we can make digital copies of ourselves.

Published on KurzweilAI.net April 7, 2001. Originally published by PsychologyToday.com February 2, 2000.

Thought to Implant 4: OnNet, please.

Hundreds of shimmering thumbnail images mist into view, spread fairly evenly across the entire field of pseudovision.

Thought: Zoom upper left, higher, into Winston's image.

Transmit: It's Nellie. Let's connect and chat over croissants. Rue des Enfants, Paris in the spring, our favorite table, yes?

Four-second pause.

Background thought: Damn it. What's taking him so long?

Receive: I'm here, ma chêre, I'm here! Let's do it!

The thumbnail field mists away, and a café scene swirls into place. Scent of honeysuckle. Paté. Wine. Light breeze. Nellie is seated at a quaint table with a plain white tablecloth. An image of Winston looking 20 and buff mists in across from her. Message thumbnails occasionally blink against the sky.

Winston: It's so good to see you again, ma chêre! It's been months! And what a gorgeous choice of bodies! The eyes are a dead giveaway, though. You always pick those raspberry eyes. Très bold, Nellita. So what's the occasion? Part of me is in the middle of a business meeting in Chicago, so I can't dally.

Nellie: Why do you always put on that muscleman body, Winston? You know how much I like your real one. Winston morphs into a man in his early 50s, still overly muscular.

Winston: (laughing) My real body? How droll! No one but my neurotechnician has seen it for years! Believe me, that's not what you want. I can do much better! He fans rapidly through a thousand images, and Nellie grimaces.

Nellie: Damn it! You're just one of Winston's MI's! Where is the real Winston? I know I used the right connection!

Winston: Nellie, I'm sorry to have to tell you this. There was a transporter accident a few weeks ago in Evanston, and well, I'm lucky they got to me in time for the full upload. I'm all of Winston that's left. The body's gone.

When Nellie contacts her friend Winston through the Internet connection in her brain, he is already, biologically speaking, dead. It is his electronic mind double, a virtual reality twin, that greets Nellie in their virtual Parisian café. What's surprising here is not so much the notion that human minds may someday live on inside computers after their bodies have expired. It's the fact that this vignette is closer at hand than most people realize. Within 30 years, the minds in those computers may just be our own.

The history of technology has shown over and over that as one mode of technology exhausts its potential, a new more sophisticated paradigm emerges to keep us moving at an exponential pace. Between 1910 and 1950, computer technology doubled in power every three years; between 1950 and 1966, it doubled every two years; and it has recently been doubling every year.

By the year 2020, your $1,000 personal computer will have the processing power of the human brain—20 million billion calculations per second (100 billion neurons times 1,000 connections per neuron times 200 calculations per second per connection). By 2030, it will take a village of human brains to match a $1,000 computer. By 2050, $1,000 worth of computing will equal the processing power of all human brains on earth.

Of course, achieving the processing power of the human brain is necessary but not sufficient for creating human level intelligence in a machine. But by 2030, we'll have the means to scan the human brain and re-create its design electronically.

Most people don't realize the revolutionary impact of that. The development of computers that match and vastly exceed the capabilities of the human brain will be no less important than the evolution of human intelligence itself some thousands of generations ago. Current predictions overlook the imminence of a world in which machines become more like humans—programmed with replicated brain synapses that re-create the ability to respond appropriately to human emotion, and humans become more like machines—our biological bodies and brains enhanced with billions of "nanobots," swarms of microscopic robots transporting us in and out of virtual reality. We have already started down this road: Human and machine have already begun to meld.

It starts with uploading, or scanning the brain into a computer. One scenario is invasive: One very thin slice at a time, scientists input a brain of choice—having been frozen just slightly before it was going to die—at an extremely high speed. This way, they can easily see every neuron, every connection and every neurotransmitter concentration represented in each synapse-thin layer.

Seven years ago, a condemned killer allowed his brain and body to be scanned in this way, and you can access all 10 billion bytes of him on the Internet. You can see for yourself every bone, muscle and section of gray matter in his body. But the scan is not yet at a high enough resolution to re-create the interneuronal connections, synapses and neurotransmitter concentrations that are the key to capturing the individuality within a human brain.

Our scanning machines today can clearly capture neural features as long as the scanner is very close to the source. Within 30 years, however, we will be able to send billions of nanobots—blood cell-size scanning machines—through every capillary of the brain to create a complete noninvasive scan of every neural feature. A shot full of nanobots will someday allow the most subtle details of our knowledge, skills and personalities to be copied into a file and stored in a computer.

We can touch and feel this technology today. We just can't make the nanobots small enough, not yet anyway. But miniaturization is another one of those accelerating technology trends. We're currently shrinking the size of technology by a factor of 5.6 per linear dimension per decade, so it is conservative to say that this scenario will be feasible in a few decades. The nanobots will capture the locations, interconnections and contents of all the nerve cell bodies, axons, dendrites, presynaptic vesicles, neurotransmitter concentrations and other relevant neural components. Using high-speed wireless communication, the nanobots will then communicate with each other and with other computers that are compiling the brain-scan database.

If this seems daunting, another scanning project, that of the human genome, was also considered ambitious when it was first introduced 12 years ago. At the time, skeptics said the task would take thousands of years, given current scanning capabilities. But the project is finishing on time nevertheless because the speed with which we can sequence DNA has grown exponentially.

Brain scanning is a prerequisite to Winston and Nellie's virtual life-and apparent immortality.

In 2029, we will swallow or inject billions of nanobots into our veins to enter a three dimensional cyberspace—a virtual reality environment. Already, neural implants are used to counteract tremors from Parkinson's disease as well as multiple sclerosis. I have a deaf friend who can now hear what I'm saying because of his cochlear implant. Under development is a retinal implant that will perform a similar function for blind people, basically replacing certain visual processing circuits of the brain. Recently, scientists from Emory University placed a chip in the brain of a paralyzed stroke victim who can now begin to communicate and control his environment directly from his brain.

But while a surgically introduced neural implant can be placed in only one or at most a few locations, nanobots can take up billions or trillions of positions throughout the brain. We already have electronic devices called neuron transistors that, noninvasively, allow communication between electronics and biological neurons. Using this technology, developed at Germany's Max Planck Institute of Biochemistry, scientists were recently able to control from their computer the movements of a living leech.

By taking up positions next to specific neurons, the nanobots will be able to detect and control their activity. For virtual reality applications, the nanobots will take up positions next to every nerve fiber coming from all five of our senses. When we want to enter a specific virtual environment, the nanobots will suppress the signals coming from our real senses and replace them with new, virtual ones. We can then cause our virtual body to move, speak and otherwise interact in the virtual environment. The nanobots would prevent our real bodies from moving; instead, we would have a virtual body in a virtual environment, which need not be the same as our real body.

Like the experiences Winston and Nellie enjoyed, this technology will enable us to have virtual interactions with other people—or simulated people—without requiring any equipment not already in our heads. And virtual reality will not be as crude as what you experience in today's arcade games. It will be as detailed and subtle as real life. So instead of just phoning a friend, you can meet in a virtual Italian bistro or stroll down a virtual tropical beach, and it will all seem real. People will be able to share any type of experience—business, social, romantic or sexual—regardless of physical proximity.

The trip to virtual reality will be readily reversible since, with your thoughts alone, you will be able to shut the nanobots off, or even direct them to leave your body. Nanobots are programmable, in that they can provide virtual reality one minute and a variety of brain extensions the next. They can change their configuration, and even alter their software.

While the combination of human-level intelligence in a machine and a computer's inherent superiority in the speed, accuracy and sharing ability of its memory will be formidable—this is not an alien invasion. It is emerging from within our human-machine civilization.

But will virtual life and its promise of immortality obviate the fear of death? Once we upload our knowledge, memories and insights into a computer, will we have acquired eternal life? First we must determine what human life is. What is consciousness anyway? If my thoughts, knowledge, experience, skills and memories achieve eternal life without me, what does that mean for me?

Consciousness—a seemingly basic tenet of "living"—is perplexing and reflects issues that have been debated since the Platonic dialogues. We assume, for instance, that other humans are conscious, but when we consider the possibility that nonhuman animals may be conscious, our understanding of consciousness is called into question.

The issue of consciousness will become even more contentious in the 21st century because nonbiological entities—read: machines—will be able to convince most of us that they are conscious. They will master all the subtle cues that we now use to determine that humans are conscious. And they will get mad if we refute their claims.

Consider this: If we scan me, for example, and record the exact state, level and position of my every neurotransmitter, synapse, neural connection and other relevant details, and then reinstantiate this massive database into a neural computer, then who is the real me? If you ask the machine, it will vehemently claim to be the original Ray. Since it will have all of my memories, it will say, "I grew up in Queens, New York, went to college at MIT, stayed in the Boston area,

sold a few artificial intelligence companies, walked into a scanner there and woke up in the machine here. Hey, this technology really works."

But there are strong arguments that this is really a different person. For one thing, old biological Ray (that's me) still exists. I'll still be here in my carbon, cell-based brain. Alas, I (the old biological Ray) will have to sit back and watch the new Ray succeed in endeavors that I could only dream of.

But New Ray will have some strong claims as well. He will say that while he is not absolutely identical to Old Ray, neither is the current version of Old Ray, since the particles making up my biological brain and body are constantly changing. It is the patterns of matter and energy that are semipermanent (that is, changing only gradually), while the actual material content changes constantly and very quickly.

Viewed in this way, my identity is rather like the pattern that water makes when rushing around a rock in a stream. The pattern remains relatively unchanged for hours, even years, while the actual material constituting the pattern-the water-is replaced in milliseconds.

This idea is consistent with the philosophical notion that we should not associate our fundamental identity with a set of particles, but rather with the pattern of matter and energy that we represent. In other words, if we change our definition of consciousness to value patterns over particles, then New Ray may have an equal claim to be the continuation of Old Ray.

One could scan my brain and reinstantiate the new Ray while I was sleeping, and I would not necessarily even know about it. If you then came to me, and said, "Good news, Ray, we've successfully reinstantiated your mind file so we won't be needing your old body and brain anymore," I may quickly realize the philosophical flaw in the argument that New Ray is a continuation of my consciousness. I may wish New Ray well, and realize that he shares my pattern, but I would nonetheless conclude that he is not me, because I'm still here.

Wherever you wind up on this debate, it is worth noting that data do not necessarily last forever. The longevity of information depends on its relevance, utility and accessibility. If you've ever tried to retrieve information from an obsolete form of data storage in an old obscure format (e.g., a reel of magnetic tape from a 1970s minicomputer), you understand the challenge of keeping software viable. But if we are diligent in maintaining our mind file, keeping current backups and porting to the latest formats and mediums, then at least a crucial aspect of who we are will attain a longevity independent of our bodies.

What does this super technological intelligence mean for the future? There will certainly be grave dangers associated with 21st century technologies. Consider unrestrained nanobot replication. The technology requires billions or trillions of nanobots in order to be useful, and the most cost-effective way to reach such levels is through self-replication, essentially the same approach used in the biological world, by bacteria, for example. So in the same way that biological self-replication gone awry (i.e., cancer) results in biological destruction, a defect in the mechanism curtailing nanobot self-replication would endanger all physical entities, biological or otherwise.

Other salient questions are: Who is controlling the nanobots? Who else might the nanobots be talking to?

Organizations, including governments, extremist groups or even a clever individual, could put trillions of undetectable nanobots in the water or food supply of an entire population. These "spy" nanobots could then monitor, influence and even control our thoughts and actions. In addition, authorized nanobots could be influenced by software viruses and other hacking techniques. Just as technology poses dangers today, there will be a panoply of risks in the decades ahead.

On a personal level, I am an optimist, and I expect that the creative and constructive applications of this technology will persevere, as I believe they do today. But there will be a valuable and increasingly vocal role for a concerned movement of Luddites—those anti-technologists inspired by early-19th-century weavers who in protest, destroyed machinery that was threatening their livelihood.

Still, I regard the freeing of the human mind from its severe physical limitations as a necessary next step in evolution. Evolution, in my view, is the purpose of life, meaning that the purpose of life-and of our lives-is to evolve.

What does it mean to evolve? Evolution moves toward greater complexity, elegance, intelligence, beauty, creativity and love. And God has been called all these things, only without any limitation, infinite. While evolution never reaches an infinite level, it advances exponentially, certainly moving in that direction. Technological evolution, therefore, moves us inexorably closer to becoming like God. And the freeing of our thinking from the severe limitations of our biological form may be regarded as an essential spiritual quest.

By the close of the next century, nonbiological intelligence will be ubiquitous. There will be few humans without some form of artificial intelligence, which is growing at a double exponential rate, whereas biological intelligence is basically at a standstill. Nonbiological thinking will be trillions of trillions of times more powerful than that of its biological progenitors, although it will be still of human origin.

Ultimately, however, the earth's technology-creating species will merge with its own computational technology. After all, what is the difference between a human brain enhanced a trillion-fold by nanobot-based implants, and a computer whose design is based on high-resolution scans of the human brain, and then extended a trillion-fold?

This may be the ominous, existential question that our own children, certainly our grandchildren, will face. But at this point, there's no turning back. And there's no slowing down.

The Coming Merging of Mind and Machine

Ray Kurzweil

http://www.kurzweilai.net/articles/art0063.html

Ray Kurzweil predicts a future with direct brain-to-computer access and conscious machines.

Published on KurzweilAI.net February 22, 2001. Originally published in Scientific American September 1, 1999.

Sometime early in the next century, the intelligence of machines will exceed that of humans. Within several decades, machines will exhibit the full range of human intellect, emotions and skills, ranging from musical and other creative aptitudes to physical movement. They will claim to have feelings and, unlike today's virtual personalities, will be very convincing when they tell us so. By 2019 a $1,000 computer will at least match the processing power of the human brain. By 2029 the software for intelligence will have been largely mastered, and the average personal computer will be equivalent to 1,000 brains.

Within three decades, the author maintains, neural implants will be available that interface directly to our brain cells. The implants would enhance sensory experiences and improve our memory and thinking.

Once computers achieve a level of intelligence comparable to that of humans, they will necessarily soar past it. For example, if I learn French, I can't readily download that learning to you. The reason is that for us, learning involves successions of stunningly complex patterns of interconnections among brain cells (neurons) and among the concentrations of biochemicals, known as neurotransmitters, that enable impulses to travel from neuron to neuron. We have no way of quickly downloading these patterns. But quick downloading will allow our nonbiological creations to share immediately what they learn with billions of other machines. Ultimately, nonbiological entities will master not only the sum total of their own knowledge but all of ours as well.

As this happens, there will no longer be a clear distinction between human and machine. We are already putting computers—neural implants—directly into people's brains to counteract Parkinson's disease and tremors from multiple sclerosis. We have cochlear implants that restore hearing. A retinal implant is being developed in the U.S. that is intended to provide at least some visual perception for some blind individuals, basically by replacing certain visual-processing circuits of the brain. Recently scientists from Emory University implanted a chip in the brain of a paralyzed stroke victim that allows him to use his brainpower to move a cursor across a computer screen.

In the 2020s neural implants will improve our sensory experiences, memory and thinking. By 2030, instead of just phoning a friend, you will be able to meet in, say, a virtual Mozambican game preserve that will seem compellingly real. You will be able to have any type of

experience—business, social, sexual—with anyone, real or simulated, regardless of physical proximity.

How Life and Technology Evolve

To gain insight into the kinds of forecasts I have just made, it is important to recognize that technology is advancing exponentially. An exponential process starts slowly, but eventually its pace increases extremely rapidly. (A fuller documentation of my argument is contained in my new book, The Age of Spiritual Machines.)

The evolution of biological life and the evolution of technology have both followed the same pattern: they take a long time to get going, but advances build on one another and progress erupts at an increasingly furious pace. We are entering that explosive part of the technological evolution curve right now.

Consider: It took billions of years for Earth to form. It took two billion more for life to begin and almost as long for molecules to organize into the first multicellular plants and animals about 700 million years ago. The pace of evolution quickened as mammals inherited Earth some 65 million years ago. With the emergence of primates, evolutionary progress was measured in mere millions of years, leading to Homo sapiens perhaps 500,000 years ago.

The evolution of technology has been a continuation of the evolutionary process that gave rise to us-the technology-creating species-in the first place. It took tens of thousands of years for our ancestors to figure out that sharpening both sides of a stone created useful tools. Then, earlier in this millennium, the time required for a major paradigm shift in technology had shrunk to hundreds of years.

The pace continued to accelerate during the 19th century, during which technological progress was equal to that of the 10 centuries that came before it. Advancement in the first two decades of the 20th century matched that of the entire 19th century. Today significant technological transformations take just a few years; for example, the World Wide Web, already a ubiquitous form of communication and commerce, did not exist just nine years ago.

Computing technology is experiencing the same exponential growth. Over the past several decades, a key factor in this expansion has been described by Moore's Law. Gordon Moore, a co-founder of Intel, noted in the mid-1960s that technologists had been doubling the density of transistors on integrated circuits every 12 months. This meant computers were periodically doubling both in capacity and in speed per unit cost. In the mid-1970s Moore revised his observation of the doubling time to a more accurate estimate of about 24 months, and that trend has persisted through the 1990s.

After decades of devoted service, Moore's Law will have run its course around 2019. By that time, transistor features will be just a few atoms in width. But new computer architectures will continue the exponential growth of computing. For example, computing cubes are already being designed that will provide thousands of layers of circuits, not just one as in today's computer chips. Other technologies that promise orders-of-magnitude increases in computing density

include nanotube circuits built from carbon atoms, optical computing, crystalline computing and molecular computing.

We can readily see the march of computing by plotting the speed (in instructions per second) per $1,000 (in constant dollars) of 49 famous calculating machines spanning the 20th century [see graph below]. The graph is a study in exponential growth: computer speed per unit cost doubled every three years between 1910 and 1950 and every two years between 1950 and 1966 and is now doubling every year. It took 90 years to achieve the first $1,000 computer capable of executing one million instructions per second (MIPS). Now we add an additional MIPS to a $1,000 computer every day.

Why Returns Accelerate

Why do we see exponential progress occurring in biological life, technology and computing? It is the result of a fundamental attribute of any evolutionary process, a phenomenon I call the Law of Accelerating Returns. As order exponentially increases (which reflects the essence of evolution), the time between salient events grows shorter. Advancement speeds up. The returns—the valuable products of the process—accelerate at a nonlinear rate. The escalating growth in the price performance of computing is one important example of such accelerating returns.

A frequent criticism of predictions is that they rely on an unjustified extrapolation of current trends, without considering the forces that may alter those trends. But an evolutionary process accelerates because it builds on past achievements, including improvements in its own means for further evolution. The resources it needs to continue exponential growth are its own increasing order and the chaos in the environment in which the evolutionary process takes place, which provides the options for further diversity. These two resources are essentially without limit.

The Law of Accelerating Returns shows that by 2019 a $1,000 personal computer will have the processing power of the human brain—20 million billion calculations per second. Neuroscientists came up with this figure by taking an estimation of the number of neurons in the brain, 100 billion, and multiplying it by 1,000 connections per neuron and 200 calculations per second per connection. By 2055, $1,000 worth of computing will equal the processing power of all human brains on Earth (of course, I may be off by a year or two).

The accelerating rate of progress in computing is demonstrated by this graph, which shows the amount of computing speed that $1,000 (in constant dollars) would buy, plotted as a function of time. Computer power per unit cost is now doubling every year.

Programming Intelligence

That's the prediction for processing power, which is a necessary but not sufficient condition for achieving human-level intelligence in machines. Of greater importance is the software of intelligence.

One approach to creating this software is to painstakingly program the rules of complex processes. We are getting good at this task in certain cases; the Cyc (as in "encyclopedia") system designed by Douglas B. Lenat of Cycorp has more than one million rules that describe the intricacies of human common sense, and it is being applied to Internet search engines so that they return smarter answers to our queries.
Another approach is "complexity theory" (also known as chaos theory) computing, in which self-organizing algorithms gradually learn patterns of information in a manner analogous to human learning. One such method, neural nets, is based on simplified mathematical models of mammalian neurons. Another method, called genetic (or evolutionary) algorithms, is based on allowing intelligent solutions to develop gradually in a simulated process of evolution.

Ultimately, however, we will learn to program intelligence by copying the best intelligent entity we can get our hands on: the human brain itself. We will reverse-engineer the human brain, and fortunately for us it's not even copyrighted!

The most immediate way to reach this goal is by destructive scanning: take a brain frozen just before it was about to expire and examine one very thin slice at a time to reveal every neuron, interneuronal connection and concentration of neurotransmitters across each gap between neurons (these gaps are called synapses). One condemned killer has already allowed his brain and body to be scanned, and all 15 billion bytes of him can be accessed on the National Library of Medicine's Web site. The resolution of these scans is not nearly high enough for our purposes, but the data at least enable us to start thinking about these issues.

We also have noninvasive scanning techniques, including high-resolution magnetic resonance imaging (MRI) and others. Their increasing resolution and speed will eventually enable us to resolve the connections between neurons. The rapid improvement is again a result of the Law of Accelerating Returns, because massive computation is the main element in higher-resolution imaging.

Another approach would be to send microscopic robots (or "nanobots") into the bloodstream and program them to explore every capillary, monitoring the brain's connections and neurotransmitter concentrations.

Fantastic Voyage

Although sophisticated robots that small are still several decades away at least, their utility for probing the innermost recesses of our bodies would be far-reaching. They would communicate wirelessly with one another and report their findings to other computers. The result would be a noninvasive scan of the brain taken from within.

Most of the technologies required for this scenario already exist, though not in the microscopic size required. Miniaturizing them to the tiny sizes needed, however, would reflect the essence of the Law of Accelerating Returns. For example, the translators on an integrated circuit have been shrinking by a factor of approximately 5.6 in each linear dimension every 10 years.

The capabilities of these embedded nanobots would not be limited to passive roles such as monitoring. Eventually they could be built to communicate directly with the neuronal circuits in our brains, enhancing or extending our mental capabilities. We already have electronic devices that can communicate with neurons by detecting their activity and either triggering nearby neurons to fire or suppressing them from firing. The embedded nanobots will be capable of reprogramming neural connections to provide virtual-reality experiences and to enhance our pattern recognition and other cognitive faculties.

To decode and understand the brain's information-processing methods (which, incidentally, combine both digital and analog methods), it is not necessary to see every connection, because there is a great deal of redundancy within each region. We are already applying insights from early stages of this reverse-engineering process. For example, in speech recognition, we have already decoded and copied the brain's early stages of sound processing.

Perhaps more interesting than this scanning-the-brain-to-understand-it approach would be scanning the brain for the purpose of downloading it. We would map the locations, interconnections, and contents of all the neurons, synapses and neurotransmitter concentrations. The entire organization, including the brain's memory, would then be re-created on a digital-analog computer.

To do this, we would need to understand local brain processes, and progress is already under way. Theodore W. Berger and his co-workers at the University of Southern California have built integrated circuits that precisely match the processing characteristics of substantial clusters of neurons. Carver A. Mead and his colleagues at the California Institute of Technology have built a variety of integrated circuits that emulate the digital-analog characteristics of mammalian neural circuits.

Developing complete maps of the human brain is not as daunting as it may sound. The Human Genome Project seemed impractical when it was first proposed. At the rate at which it was possible to scan genetic codes 12 years ago, it would have taken thousands of years to complete the genome. But in accordance with the Law of Accelerating Returns, the ability to sequence DNA has been accelerating. The latest estimates are that the entire human genome will be completed in just a few years.

By the third decade of the 21st century, we will be in a position to create complete, detailed maps of the computationally relevant features of the human brain and to re-create these designs in advanced neural computers. We will provide a variety of bodies for our machines, too, from virtual bodies in virtual reality to bodies comprising swarms of nanobots. In fact, humanoid robots that ambulate and have lifelike facial expressions are already being developed at several laboratories in Tokyo.

Will It Be Conscious?

Such possibilities prompt a host of intriguing issues and questions. Suppose we scan someone's brain and reinstate the resulting "mind file" into a suitable computing medium. Will the entity that emerges from such an operation be conscious? This being would appear to others to have

very much the same personality, history and memory. For some, that is enough to define consciousness. For others, such as physicist and author James Trefil, no logical reconstruction can attain human consciousness, although Trefil concedes that computers may become conscious in some new way.
At what point do we consider an entity to be conscious, to be self-aware, to have free will? How do we distinguish a process that is conscious from one that just acts as if it is conscious? If the entity is very convincing when it says, "I'm lonely, please keep me company," does that settle the issue?

If you ask the "person" in the machine, it will strenuously claim to be the original person. If we scan, let's say, me and reinstate that information into a neural computer, the person who emerges will think he is (and has been) me (or at least he will act that way). He will say, "I grew up in Queens, New York, went to college at M.I.T., stayed in the Boston area, walked into a scanner there and woke up in the machine here. Hey, this technology really works." But wait, is this really me? For one thing, old Ray (that's me) still exists in my carbon-cell-based brain.

Will the new entity be capable of spiritual experiences? Because its brain processes are effectively identical, its behavior will be comparable to that of the person it is based on. So it will certainly claim to have the full range of emotional and spiritual experiences that a person claims to have.

No objective test can absolutely determine consciousness. We cannot objectively measure subjective experience (this has to do with the very nature of the concepts "objective" and "subjective"). We can measure only correlates of it, such as behavior. The new entities will appear to be conscious, and whether or not they actually are will not affect their behavior. Just as we debate today the consciousness of nonhuman entities such as animals, we will surely debate the potential consciousness of nonbiological intelligent entities. From a practical perspective, we will accept their claims. They'll get mad if we don't.

Before the next century is over, the Law of Accelerating Returns tells us, Earth's technology-creating species-us-will merge with our own technology. And when that happens, we might ask: What is the difference between a human brain enhanced a millionfold by neural implants and a nonbiological intelligence based on the reverse-engineering of the human brain that is subsequently enhanced and expanded?

The engine of evolution used its innovation from one period (humans) to create the next (intelligent machines). The subsequent milestone will be for the machines to create their own next generation without human intervention.

An evolutionary process accelerates because it builds on its own means for further evolution. Humans have beaten evolution. We are creating intelligent entities in considerably less time than it took the evolutionary process that created us. Human intelligence—a product of evolution—has transcended it. So, too, the intelligence that we are now creating in computers will soon exceed the intelligence of its creators.

Visions of the Future

Science fiction becoming fact: instant information everywhere, virtually infinite bandwidth, implanted computer, nanotechnology breakthroughs. What's next?

The Matrix Loses Its Way: Reflections on 'Matrix' and 'Matrix Reloaded'

Ray Kurzweil

http://www.kurzweilai.net/articles/art0580.html

The Matrix Reloaded is crippled by senseless fighting and chase scenes, weak plot and character development, tepid acting, and sophomoric dialogues. It shares the dystopian, Luddite perspective of the original movie, but loses the elegance, style, originality, and evocative philosophical musings of the original.

Published on KurzweilAI.net May 18, 2003

You're going to love Matrix Reloaded-that is, if you're a fan of endless Kung Fu fights, repetitive chase scenes, a meandering and poorly paced plot, and sophomoric philosophical musings. For much of its 2 hours and 18 minutes, I felt like I was stuck looking over the shoulder of a ten-year-old playing a video game.

It's too bad, because the original Matrix was a breakout film, introducing audiences to a new approach to movie making, while reflecting in an elegant way on pivotal ideas about the future. Although I disagree with its essentially Luddite stance, it raised compelling issues that have drawn intense reactions, including thousands of articles and at least a half dozen books

Is Matrix-style VR feasible?

There is a lot more to say about the original Matrix than this derivative and overwrought sequel, so let me start with that. The Matrix introduced its vast audience to the idea of full-immersion virtual reality, to what Morpheus (Laurence Fishburne) describes as a "neural interactive simulation" that is indistinguishable from real reality. I have been asked many times whether virtual reality with this level of realism will be feasible and when.

As I described in my chapter "The Human Machine Merger: Are We Heading for The Matrix?" in the book *Taking the Red Pill*[1], virtual reality will become a profoundly transforming technology by 2030. By then, nanobots (robots the size of human blood cells or smaller, built with key features at the multi-nanometer-billionth of a meter-scale) will provide fully immersive, totally convincing virtual reality in the following way. The nanobots take up positions in close physical proximity to every interneuronal connection coming from all of our senses (e.g., eyes, ears, skin). We already have the technology for electronic devices to communicate with neurons in both directions that requires no direct physical contact with the neurons.

[1] Glenn Yeffeth, Ed., Taking the Red Pill: Science, Philosophy and Religion in The Matrix (Ben Bella Books, April 2003)

For example, scientists at the Max Planck Institute have developed "neuron transistors" that can detect the firing of a nearby neuron, or alternatively, can cause a nearby neuron to fire, or suppress it from firing. This amounts to two-way communication between neurons and the electronic-based neuron transistors. The Institute scientists demonstrated their invention by controlling the movement of a living leech from their computer. Nanobot-based virtual reality is not yet feasible in size and cost, but we have made a good start in understanding the encoding of sensory signals. For example, Lloyd Watts and his colleagues have developed a detailed model of the sensory coding and transformations that take place in the auditory processing regions of the human brain. We are at an even earlier stage in understanding the complex feedback loops and neural pathways in the visual system.

When we want to experience real reality, the nanobots just stay in position (in the capillaries) and do nothing. If we want to enter virtual reality, they suppress all of the inputs coming from the real senses, and replace them with the signals that would be appropriate for the virtual environment. You (i.e., your brain) could decide to cause your muscles and limbs to move as you normally would, but the nanobots again intercept these interneuronal signals, suppress your real limbs from moving, and instead cause your virtual limbs to move and provide the appropriate movement and reorientation in the virtual environment.

The Web will provide a panoply of virtual environments to explore. Some will be recreations of real places, others will be fanciful environments that have no "real" counterpart. Some indeed would be impossible in the physical world (perhaps because they violate the laws of physics). We will be able to "go" to these virtual environments by ourselves, or we will meet other people there, both real and virtual people.

By 2030, going to a web site will mean entering a full-immersion virtual-reality environment. In addition to encompassing all of the senses, these shared environments could include emotional overlays, since the nanobots will be capable of triggering the neurological correlates of emotions, sexual pleasure, and other derivatives of our sensory experience and mental reactions.

The portrayal of virtual reality in the Matrix is a bit more primitive than this. The use of bioports in the back of the neck reflects a lack of imagination on how full-immersion virtual reality from within the nervous system is likely to work. The idea of a plug is an old fashioned notion that we are already starting to get away from in our machines. By the time the Matrix is feasible, we will have far more elegant means of wirelessly accessing the human nervous system from within.

Virtual reality, as conceived of in the Matrix, is evil. Morpheus describes the Matrix as "a computer-generated dream world to keep us under control." We saw similar portrayals of the Internet prior to its creation. Early fiction, such as the novels 1984 and *Brave New World*, portrayed the worldwide communications network as essentially evil, a means for totalitarian control of humankind. Now that we actually have a worldwide communications network, we can see that the reality has turned out rather different.

Like any technology, the Internet empowers both our creative and destructive inclinations, but overall the advent of worldwide decentralized electronic communication has been a powerful democratizing force. It was not Yeltsin standing on a tank that overthrew Soviet control during

the 1991 revolt after the coup against Gorbachev. Rather it was the early forms of electronic messaging (such as fax machines and an early form of email based on teletype machines), forerunners to the Internet, that prevented the totalitarian forces from keeping the public in the dark. We can trace the movement towards democracy throughout the 1990s to the emergence of this worldwide communications network.

In my view, the advent of virtual reality will reflect a similar amplification of creative human communication. We have one form of virtual reality already. It's called the telephone, and it is a way to "be together" even if physically apart, at least as far as the auditory sense is concerned. When we add all of the other senses to virtual reality, it will be a similar strengthening of human communication.

A Dystopian, Luddite Perspective

The dystopian, Luddite perspective of the Wachowski brothers can be seen in its view of the birth of artificial intelligence as the source of all evil. In one of Morpheus' "sermons," he tells Neo (Keanu Reeves) that "in the early 21st century, all of mankind united and marveled at our magnificence as we gave birth to AI [artificial intelligence], a singular construction that spawned an entire race of machines." Morpheus goes on to explain how this singular construction became a runaway phenomenon as it reproduced itself and ultimately enslaved humankind.

The movie celebrates those humans who choose to be completely unaltered by technology, even spurning the bioport. Incidentally, in my book *The Age of Spiritual Machines*[2], I refer to such people as MOSHs (Mostly Original Substrate Humans). The movie's position reflects a growing sentiment in today's world to maintain a distinct separation of the natural- and human-created worlds. The reality, however, is that these worlds are rapidly merging. We already have a variety of neural implants that are repairing human brains afflicted by disease or disability, for example, an FDA-approved neural implant that replaces the region of neurons destroyed by Parkinson's Disease, cochlear implants for the deaf, and emerging retinal implants for the blind.

My view is that the prospect of "strong AI" (AI at or beyond human intelligence) will serve to amplify human civilization much the same way that our technology does today. As a society, we routinely accomplish intellectual achievements that would be impossible without the level of computer intelligence we already have. Ultimately, we will merge our own biological intelligence with our own creations as a way of continuing the exponential expansion of human knowledge and creative potential.

However, I do not completely reject the specter of AI turning on its creators, as portrayed in the Matrix. It is a possible downside scenario, what Nick Bostrom calls an "existential risk[3]." There has been a great deal of discussion recently about future dangers that Bill Joy [4],[5],[6] has labeled

---

[2] Ray Kurzweil, *The Age of Spiritual Machines*, Penguin USA, 1999
[3] Nick Bostrom, "Existential Risks: Analyzing Human Extinction Scenario and Related Hazards," 2001
[4] Bill Joy, "Why the future doesn't need us," Wired, April 2000
[5] Ray Kurzweil, In Response to," KurzweilAI.net July 25, 2001
[6] Ray Kurzweil, "Testimony of Ray Kurzweil on the Societal Implications of Nanotechnology," KurzweilAI.net, April 9, 2003

"GNR" (genetics, nanotechnology, and robotics). The "G" peril, which is the destructive potential of bioengineered pathogens, is the danger we are now struggling with. Our first defense from "G" will need to be more "G," for example bioengineered antiviral medications.

Ultimately, we will provide a true defense from "G" by using "N," nanoengineered entities that are smaller, faster, and smarter than mere biological entities. However, the advent of fully realized nanotechnology will introduce a new set of profound dangers. Our defense from "N" will also initially be created from defensive nanotechnology, but the ultimate defense from "N" will be "R," small robots that are intelligent at human levels and beyond, in other words, strong AI. But then the question arises: what will defend us from malevolent AI? The only possible answer is "friendly AI[7]."

Unfortunately there is nothing we can do today to assure that AI will be friendly. Based on this, some observers such as Bill Joy call for us to relinquish the pursuit of these technologies. The reality, however, is that such relinquishment is not possible without instituting a totalitarian government that bans all of technology (which is the essential theme of Brave New World). It's the same story with human intelligence. The only defense we have had throughout human history from malevolent human intelligence is for more enlightened human intelligence to confront its more deviant forms. Our imperfect record in accomplishing this is at least one key reason that there is so much concern with GNR.

Glitches

There are problems and inconsistencies with the conception of virtual reality in the Matrix. The most obvious is the absurd notion of the machines keeping all of the humans alive to use them as energy sources. Humans are capable of many things, but being an effective battery is not one of them. Our biological bodies do not generate any significant levels of useful energy. Moreover, we require more energy than we produce. Morpheus acknowledges that the machines needed more than just humans for energy when he tells Neo "25,000 BTU of body heat combined with a form of fusion [provide] the machines all the energy they need." But if the machines have fusion technology, then they clearly would not need humans.

In his chapter "Glitches in The Matrix. . ..And How to Fix Them," (also in the book *Taking the Red Pill*) Peter Lloyd surmises that "the machines are harnessing the spare brainpower of the human race as a colossal distributed processor for controlling the nuclear fusion reactions." This is a creative fix, but equally unfounded. Human brains are not an attractive building block for a distributed processor. The electrochemical signaling pathway in the human brain is extremely slow: about 200 calculations per second, which is at least 10 million times slower than today's electronics. The architecture of our brains is relatively fixed and unsuitable for harnessing into a parallel network. Moreover, the human brains in the story are presumably being actively used to guide the human lives in the virtual Matrix world. If the AI's in the matrix are smart enough to create fusion power, they would not need a network of human brains to control it.

[7] Eliezer S. Yudkowsky, "What is Friendly AI?," KurzweilAI.net, May 3, 2001

There are other absurdities, such as the requirement to find an old fashioned "land line" (telephone) to exit the Matrix. Lloyd provides a creative rationalization for this also (the land lines have fixed network addresses in the Matrix operating system that the Nebuchadnezzar's computer can access), but given the inherent flexibility in a virtual reality environment, it is clear that the reason for this requirement has more to do with the Wachowski brothers' desire to celebrate old-fashioned technology as embodying human values.

There are many arbitrary rules and limitations in the Matrix that don't make sense. Why bother fighting the agents at all (other than for the obvious "Kung Fu" cinematic reasons) when they cannot be destroyed? Why not just run away, or in the new movie, fly away?

Another attractive feature of the original Matrix movie was its philosophical musings, albeit a hodge podge of metaphorical allusions. There's Neo as the Christian Messiah who returns to deliver humanity from evil. There's the Buddhist notion that everything we see, hear and touch is an illusion. Of course, one might point out that the true reality in the Matrix is a lot grimier and grimmer than the Buddhist idea of enlightenment. We hear the martial arts philosophy (borrowed from Star Wars) of freeing yourself from rational thinking to let one's inner warrior emerge.

Then there is the green philosophy of humanity as inimical to its natural environment. This view is actually articulated by Agent Smith, who describes humanity as "a virus that does not maintain equilibrium with its environment." Most of all, we are treated to a Luddite celebration of pure humanity, along with the 19th century and early 20th century technologies of rotary phones and old gear boxes, which presumably reflect human purity.

My overall reaction to this conception is that the human rebels will need advanced technology at least comparable to that of the evil AI's if they are to prevail. The film's notion that advanced technology is inherently evil is misplaced. Technology is power, and whoever has its power will prevail. The "machines" as portrayed in the Matrix do appear to be malevolent, but the rebels are not likely to survive with their old fashioned gear boxes. However, with the script in the hands of the Wachowski brothers, we can assume that the Rebels will nonetheless have a fighting chance.

Matrix Reloaded

Which brings us to The Matrix Reloaded. Like Star Wars and Alien, also breakout movies in their time, this sequel loses the elegance, style, and originality of the original. The new film wallows in endless battle and chase scenes. Moreover, these confrontations lack any real dramatic tension. The producers are constantly changing the rules of engagement so one never thinks, "how are they going to get out of this jam?" One has only the sense that a particular character will continue if the Wachowski brothers want that character around for their own cinematic reasons. They are continually coming up with arbitrary new rules and exceptions to the rules.

Much of the fighting makes little sense. Given that the evil twin apparitions are able to magically transport themselves directly into Trinity's vehicle, and Neo is able to fly like Superman, the hand to hand combat and use of knives and poles lacks even the logic of a video game. For that matter, the two scenes of Neo battling the 100 Smiths looked exactly like a video game. Like so

much of the action, these scenes seemed superfluous and time wasting. Smith is no longer an agent, and plays no clear role in the story, to the extent that there was any attempt to tell a coherent story.

About two thirds of the way through this sequel, I turned to my companion and asked "whatever happened to the plot, wasn't there something about 250,000 Sentinels attacking Zion, the last human city?" My companion responded that it seemed that "plot" was a four letter word to the movie makers. Of course, there wasn't much time for plot development, given all of the devotion to chasing and fighting, not to mention an equally drawn out gratuitous sex scene (well, at least there is one reason to go see this film.

If plot development was weak, character development was worse. Many reviewers of the first Matrix movie noted that Keanu Reeves could not act. But his acting in the first Matrix is downright Shakespearian compared to the sequel. At least in the original, there was some portrayal of Neo's struggle with his discovery of the true nature of the Matrix, of his grappling with his role as "the one," and his coming-of-age tutorials.

In Reloaded, Reeves acts like he's had a lobotomy, sleepwalking or rather sleep-flying through the whole movie. His lover, Trinity (Carrie-Anne Moss), is equally distant and unemotional, acting like a frustrated librarian with a black belt. Morpheus was appealing in the first movie with his earnest confidence and wisdom. In the new film, he's like a preacher on morphine, which quickly gets tiresome.

The philosophical dialogues, which were refreshing in the original, sound like late-night college banter in the sequel. As for the technology of the movie itself, there was really nothing special here. They did trash about 100 General Motors cars on a multi-million dollar roadway built especially for the movie, but aside from bigger explosions, the effects were the opposite of riveting. Some of the organic backgrounds of the city of Zion were attractive, but they were all illustrated, and lacked the genuine warmth of a real human environment, which the movie professes to celebrate. The Wachowski brothers' notion of human celebration is also a bit weird as portrayed in the retro rave festivities on Zion to honor the return of the rebels.

Although I take issue with the strong Luddite posture of the original Matrix, I recognized its importance as a forceful and stylish articulation in cinematic terms of salient 21st century issues. Unfortunately, the sequel throws away this metaphysical mantle.

Reflections on Stephen Wolfram's 'A New Kind of Science'

Ray Kurzweil

http://www.kurzweilai.net/articles/art0464.html

In his remarkable new book, Stephen Wolfram asserts that cellular automata operations underlie much of the real world. He even asserts that the entire Universe itself is a big cellular-automaton computer. But Ray Kurzweil challenges the ability of these ideas to fully explain the complexities of life, intelligence, and physical phenomena.

Published on KurzweilAI.net May 13, 2002

Stephen Wolfram's *A New Kind of Science* is an unusually wide-ranging book covering issues basic to biology, physics, perception, computation, and philosophy. It is also a remarkably narrow book in that its 1,200 pages discuss a singular subject, that of cellular automata. Actually, the book is even narrower than that. It is principally about cellular automata rule 110 (and three other rules which are equivalent to rule 110), and its implications.

It's hard to know where to begin in reviewing Wolfram's treatise, so I'll start with Wolfram's apparent hubris, evidenced in the title itself. A new science would be bold enough, but Wolfram is presenting a new *kind* of science, one that should change our thinking about the whole enterprise of science. As Wolfram states in chapter 1, "I have come to view [my discovery] as one of the more important single discoveries in the whole history of theoretical science."[1]

This is not the modesty that we have come to expect from scientists, and I suspect that it may earn him resistance in some quarters. Personally, I find Wolfram's enthusiasm for his own ideas refreshing. I am reminded of a comment made by the Buddhist teacher Guru Amrit Desai, when he looked out of his car window and saw that he was in the midst of a gang of Hell's Angels. After studying them in great detail for a long while, he finally exclaimed, "They really love their motorcycles." There was no disdain in this observation. Guru Desai was truly moved by the purity of their love for the beauty and power of something that was outside themselves.

Well, Wolfram really loves his cellular automata. So much so, that he has immersed himself for over ten years in the subject and produced what can only be regarded as a tour de force on their mathematical properties and potential links to a broad array of other endeavors. In the end notes, which are as extensive as the book itself, Wolfram explains his approach: "There is a common style of understated scientific writing to which I was once a devoted subscriber. But at some point I discovered that more significant results are usually incomprehensible if presented in this style…. And so in writing this book I have chosen to explain straightforwardly the importance I believe my various results have."[2] Perhaps Wolfram's successful technology business career may also have had its influence here, as entrepreneurs are rarely shy about articulating the benefits of their discoveries.

So what is the discovery that has so excited Wolfram? As I noted above, it is cellular automata rule 110, and its behavior. There are some other interesting automata rules, but rule 110 makes

the point well enough. A cellular automaton is a simple computational mechanism that, for example, changes the color of each cell on a grid based on the color of adjacent (or nearby) cells according to a transformation rule. Most of Wolfram's analyses deal with the simplest possible cellular automata, specifically those that involve just a one-dimensional line of cells, two possible colors (black and white), and rules based only on the two immediately adjacent cells. For each transformation, the color of a cell depends only on its own previous color and that of the cell on the left and the cell on the right. Thus there are eight possible input situations (i.e., three combinations of two colors). Each rule maps all combinations of these eight input situations to an output (black or white). So there are $2^8 = 256$ possible rules for such a one-dimensional, two-color, adjacent-cell automaton. Half of the 256 possible rules map onto the other half because of left-right symmetry. We can map half of them again because of black-white equivalence, so we are left with 64 rule types. Wolfram illustrates the action of these automata with two-dimensional patterns in which each line (along the Y axis) represents a subsequent generation of applying the rule to each cell in that line.

Most of the rules are degenerate, meaning they create repetitive patterns of no interest, such as cells of a single color, or a checkerboard pattern. Wolfram calls these rules Class 1 automata. Some rules produce arbitrarily spaced streaks that remain stable, and Wolfram classifies these as belonging to Class 2. Class 3 rules are a bit more interesting in that recognizable features (e.g., triangles) appear in the resulting pattern in an essentially random order. However, it was the Class 4 automata that created the "ah ha" experience that resulted in Wolfram's decade of devotion to the topic. The Class 4 automata, of which Rule 110 is the quintessential example, produce surprisingly complex patterns that do not repeat themselves. We see artifacts such as lines at various angles, aggregations of triangles, and other interesting configurations. The resulting pattern is neither regular nor completely random. It appears to have some order, but is never predictable.

Why is this important or interesting? Keep in mind that we started with the simplest possible starting point: a single black cell. The process involves repetitive application of a very simple rule[3]. From such a repetitive and deterministic process, one would expect repetitive and predictable behavior. There are two surprising results here. One is that the results produce apparent randomness. Applying every statistical test for randomness that Wolfram could muster, the results are completely unpredictable, and remain (through any number of iterations) effectively random. However, the results are more interesting than pure randomness, which itself would become boring very quickly. There are discernible and interesting features in the designs produced, so the pattern has some order and apparent intelligence. Wolfram shows us many examples of these images, many of which are rather lovely to look at.

Wolfram makes the following point repeatedly: "Whenever a phenomenon is encountered that seems complex it is taken almost for granted that the phenomenon must be the result of some underlying mechanism that is itself complex. But my discovery that simple programs can produce great complexity makes it clear that this is not in fact correct."[4]

I do find the behavior of Rule 110 rather delightful. However, I am not entirely surprised by the idea that simple mechanisms can produce results more complicated than their starting conditions. We've seen this phenomenon in fractals (i.e., repetitive application of a simple transformation

rule on an image), chaos and complexity theory (i.e., the complex behavior derived from a large number of agents, each of which follows simple rules, an area of study that Wolfram himself has made major contributions to), and self-organizing systems (e.g., neural nets, Markov models), which start with simple networks but organize themselves to produce apparently intelligent behavior. At a different level, we see it in the human brain itself, which starts with only 12 million bytes of specification in the genome, yet ends up with a complexity that is millions of times greater than its initial specification[5].

It is also not surprising that a deterministic process can produce apparently random results. We have had random number generators (e.g., the "randomize" function in Wolfram's program "Mathematica") that use deterministic processes to produce sequences that pass statistical tests for randomness. These programs go back to the earliest days of computer software, e.g., early versions of Fortran. However, Wolfram does provide a thorough theoretical foundation for this observation.

Wolfram goes on to describe how simple computational mechanisms can exist in nature at different levels, and that these simple and deterministic mechanisms can produce all of the complexity that we see and experience. He provides a myriad of examples, such as the pleasing designs of pigmentation on animals, the shape and markings of shells, and the patterns of turbulence (e.g., smoke in the air). He makes the point that computation is essentially simple and ubiquitous. Since the repetitive application of simple computational transformations can cause very complex phenomena, as we see with the application of Rule 110, this, according to Wolfram, is the true source of complexity in the world.

My own view is that this is only partly correct. I agree with Wolfram that computation is all around us, and that some of the patterns we see are created by the equivalent of cellular automata. But a key issue is to ask is this: *Just how complex are the results of Class 4 Automata*?

Wolfram effectively sidesteps the issue of degrees of complexity. There is no debate that a degenerate pattern such as a chessboard has no effective complexity. Wolfram also acknowledges that mere randomness does not represent complexity either, because pure randomness also becomes predictable in its pure lack of predictability. It is true that the interesting features of a Class 4 automata are neither repeating nor pure randomness, so I would agree that they are more complex than the results produced by other classes of Automata. *However*, there is nonetheless a distinct limit to the complexity produced by these Class 4 automata. The many images of Class 4 automata in the book all have a similar look to them, and although they are non-repeating, they are interesting (and intelligent) only to a degree. Moreover, they do not continue to evolve into anything more complex, nor do they develop new types of features. One could run these automata for trillions or even trillions of trillions of iterations, and the image would remain at the same limited level of complexity. They do not evolve into, say, insects, or humans, or Chopin preludes, or anything else that we might consider of a higher order of complexity than the streaks and intermingling triangles that we see in these images.

Complexity is a continuum. In the past, I've used the word "order" as a synonym for complexity, which I have attempted to define as "information that fits a purpose."[6] A completely predictable process has zero order. A high level of information alone does not necessarily imply a high level

of order either. A phone book has a lot of information, but the level of order of that information is quite low. A random sequence is essentially pure information (since it is not predictable), but has no order. The output of Class 4 automata does possess a certain level of order, and they do survive like other persisting patterns. But the pattern represented by a human being has a far higher level of order or complexity. Human beings fulfill a highly demanding purpose in that they survive in a challenging ecological niche. Human beings represent an extremely intricate and elaborate hierarchy of other patterns. Wolfram regards any pattern that combines some recognizable features and unpredictable elements to be effectively equivalent to one another, but he does not show how a Class 4 automaton can ever increase its complexity, let alone to become a pattern as complex as a human being.

There is a missing link here in how one gets from the interesting, but ultimately routine patterns of a cellular automaton to the complexity of persisting structures that demonstrate higher levels of intelligence. For example, these class 4 patterns are not capable of solving interesting problems, and no amount of iteration moves them closer to doing so. Wolfram would counter that a rule 110 automaton could be used as a "universal computer."[7] However, by itself a universal computer is not capable of solving intelligent problems without what I would call "software." It is the complexity of the software that runs on a universal computer that is precisely the issue.

One might point out that the Class 4 patterns I'm referring to result from the simplest possible cellular automata (i.e., one-dimensional, two-color, two-neighbor rules). What happens if we increase the dimensionality, e.g., go to multiple colors, or even generalize these discrete cellular automata to continuous functions? Wolfram addresses all of this quite thoroughly. The results produced from more complex automata are essentially the same as those of the very simple ones. We obtain the same sorts of interesting but ultimately quite limited patterns. Wolfram makes the interesting point that we do not need to use more complex rules to get the complexity (of Class 4 automata) in the end result. But I would make the converse point that we are unable to increase the complexity of the end result through either more complex rules or through further iteration. So cellular automata only get us so far.

So how do we get from these interesting but limited patterns of Class 4 automata to those of insects, or humans or Chopin preludes? One concept we need to add is conflict, i.e., evolution. If we add another simple concept to that of Wolfram's simple cellular automata, i.e., an evolutionary algorithm, we start to get far more interesting, and more intelligent results. Wolfram would say that the Class 4 automata and an evolutionary algorithm are "computationally equivalent." But that is only true on what I could regard as the "hardware" level. On the software level, the order of the patterns produced are clearly different, and of a different order of complexity.

An evolutionary algorithm can start with randomly generated potential solutions to a problem. The solutions are encoded in a digital genetic code. We then have the solutions compete with each other in a simulated evolutionary battle. The better solutions survive and procreate in a simulated sexual reproduction in which offspring solutions are created, drawing their genetic code (i.e., encoded solutions) from two parents. We can also introduce a rate of genetic mutation. Various high-level parameters of this process, such as the rate of mutation, the rate of offspring,

etc., are appropriately called "God parameters" and it is the job of the engineer designing the evolutionary algorithm to set them to reasonably optimal values. The process is run for many thousands of generations of simulated evolution, and at the end of the process, one is likely to find solutions that are of a distinctly higher order than the starting conditions. The results of these evolutionary (sometimes called genetic) algorithms can be elegant, beautiful, and intelligent solutions to complex problems. They have been used, for example, to create artistic designs, designs for artificial life forms in artificial life experiments, as well as for a wide range of practical assignments such as designing jet engines. Genetic algorithms are one approach to "narrow" artificial intelligence, that is, creating systems that can perform specific functions that used to require the application of human intelligence.

But something is still missing. Although genetic algorithms are a useful tool in solving specific problems, they have never achieved anything resembling "strong AI," i.e., aptitude resembling the broad, deep, and subtle features of human intelligence, particularly its powers of pattern recognition and command of language. Is the problem that we are not running the evolutionary algorithms long enough? After all, humans evolved through an evolutionary process that took billions of years. Perhaps we cannot recreate that process with just a few days or weeks or computer simulation. However, conventional genetic algorithms reach an asymptote in their level of performance, so running them for a longer period of time won't help.

A third level (beyond the ability of cellular processes to produce apparent randomness and genetic algorithms to produce focused intelligent solutions) is to perform evolution on multiple levels. Conventional genetic algorithms only allow evolution within the narrow confines of a narrow problem, and a single means of evolution. The genetic code itself needs to evolve; the rules of evolution need to evolve. Nature did not stay with a single chromosome, for example. There have been many levels of indirection incorporated in the natural evolutionary process. And we require a complex environment in which evolution takes place.

To build strong AI, we will short circuit this process, however, by reverse engineering the human brain, a project well under way, thereby benefiting from the evolutionary process that has already taken place. We will be applying evolutionary algorithms within these solutions just as the human brain does. For example, the fetal wiring is initially random in certain regions, with the majority of connections subsequently being destroyed during the early stages of brain maturation as the brain self-organizes to make sense of its environment and situation.

But back to cellular automata. Wolfram applies his key insight, which he states repeatedly, that we obtain surprisingly complex behavior from the repeated application of simple computational transformations - to biology, physics, perception, computation, mathematics, and philosophy. Let's start with biology.

Wolfram writes, "Biological systems are often cited as supreme examples of complexity in nature, and it is not uncommon for it to be assumed that their complexity must be somehow of a fundamentally higher order than other systems. . . . What I have come to believe is that many of the most obvious examples of complexity in biological systems actually have very little to do with adaptation or natural selection. And instead . . . they are mainly just another consequence of

the very basic phenomenon that I have discovered. . . .that in almost any kind of system many choices of underlying rules inevitably lead to behavior of great complexity."[8]

I agree with Wolfram that some of what passes for complexity in nature is the result of cellular-automata type computational processes. However, I disagree with two fundamental points. First, the behavior of a Class 4 automaton, as the many illustrations in the book depict, do not represent "behavior of great complexity." It is true that these images have a great deal of unpredictability (i.e., randomness). It is also true that they are not just random but have identifiable features. But the complexity is fairly modest. And this complexity never evolves into patterns that are at all more sophisticated.

Wolfram considers the complexity of a human to be equivalent to that a Class 4 automaton because they are, in his terminology, "computationally equivalent." But class 4 automata and humans are only computational equivalent in the sense that any two computer programs are computationally equivalent, i.e., both can be run on a Universal Turing machine. It is true that computation is a universal concept, and that all software is equivalent on the hardware level (i.e., with regard to the nature of computation), but it is not the case that all software is of the same order of complexity. The order of complexity of a human is greater than the interesting but ultimately repetitive (albeit random) patterns of a Class 4 automaton.

I also disagree that the order of complexity that we see in natural organisms is not a primary result of "adaptation or natural selection." The phenomenon of randomness readily produced by cellular automaton processes is a good model for fluid turbulence, but not for the intricate hierarchy of features in higher organisms. The fact that we have phenomena greater than just the interesting but fleeting patterns of fluid turbulence (e.g., smoke in the wind) in the world is precisely the result of the chaotic crucible of conflict over limited resources known as evolution.

To be fair, Wolfram does not negate adaptation or natural selection, but he over-generalizes the limited power of complexity resulting from simple computational processes. When Wolfram writes, "in almost any kind of system many choices of underlying rules inevitably lead to behavior of great complexity," he is mistaking the random placement of simple features that result from cellular processes for the true complexity that has resulted from eons of evolution.

Wolfram makes the valid point that certain (indeed most) computational processes are not predictable. In other words, we cannot predict future states without running the entire process. I agree with Wolfram that we can only know the answer in advance if somehow we can simulate a process at a faster speed. Given that the Universe runs at the fastest speed it can run, there is usually no way to short circuit the process. However, we have the benefits of the mill of billions of years of evolution, which is responsible for the greatly increased order of complexity in the natural world. We can now benefit from it by using our evolved tools to reverse-engineer the products of biological evolution.

Yes, it is true that some phenomena in nature that may appear complex at some level are simply the result of simple underlying computational mechanisms that are essentially cellular automata at work. The interesting pattern of triangles on a "tent olive" shell or the intricate and varied patterns of a snowflake are good examples. I don't think this is a new observation, in that we've

always regarded the design of snowflakes to derive from a simple molecular computation-like building process. However, Wolfram does provide us with a compelling theoretical foundation for expressing these processes and their resulting patterns. But there is more to biology than Class 4 patterns.

I do appreciate Wolfram's strong argument, however, that nature is not as complex as it often appears to be. Some of the key features of the paradigm of biological systems, which differ from much of our contemporary designed technology, are that it is massively parallel, and that apparently complex behavior can result from the intermingling of a vast number of simpler systems. One example that comes to mind is Marvin Minsky's theory of intelligence as a "Society of Mind" in which intelligence may result from a hierarchy of simpler intelligences with simple agents not unlike cellular automata at the base.

However, cellular automata on their own do not evolve sufficiently. They quickly reach a limited asymptote in their order of complexity. An evolutionary process involving conflict and competition is needed.

For me, the most interesting part of the book is Wolfram's thorough treatment of computation as a simple and ubiquitous phenomenon. Of course, we've known for over a century that computation is inherently simple, i.e., we can build any possible level of complexity from a foundation of the simplest possible manipulations of information.

For example, Babbage's computer provided only a handful of operation codes, yet provided (within its memory capacity and speed) the same kinds of transformations as do modern computers. The complexity of Babbage's invention stemmed only from the details of its design, which indeed proved too difficult for Babbage to implement using the 19th century mechanical technology available to him.

The "Turing Machine," Alan Turing's theoretical conception of a universal computer in 1950, provides only 7 very basic commands[9], yet can be organized to perform any possible computation. The existence of a "Universal Turing Machine," which can simulate any possible Turing Machine (that is described on its tape memory), is a further demonstration of the universality (and simplicity) of computation. In what is perhaps the most impressive analysis in his book, Wolfram shows how a Turing Machine with only two states and five possible colors can be a Universal Turing Machine. For forty years, we've thought that a Universal Turing Machine had to be more complex than this[10]. Also impressive is Wolfram's demonstration that Cellular Automaton Rule 110 is capable of universal computation (given the right software).

In my 1990 book, I showed how any computer could be constructed from "a suitable number of [a] very simple device," namely the "nor" gate[11]. This is not exactly the same demonstration as a universal Turing machine, but it does demonstrate that any computation can be performed by a cascade of this very simple device (which is simpler than Rule 110), given the right software (which would include the connection description of the nor gates).[12]

The most controversial thesis in Wolfram's book is likely to be his treatment of physics, in which he postulates that the Universe is a big cellular-automaton computer. Wolfram is hypothesizing

that there is a digital basis to the apparently analog phenomena and formulas in physics, and that we can model our understanding of physics as the simple transformations of a cellular automaton.

Others have postulated this possibility. Richard Feynman wondered about it in considering the relationship of information to matter and energy. Norbert Weiner heralded a fundamental change in focus from energy to information in his 1948 book Cybernetics, and suggested that the transformation of information, not energy, was the fundamental building block for the Universe.

Perhaps the most enthusiastic proponent of an information-based theory of physics was Edward Fredkin, who in the early 1980s proposed what he called a new theory of physics based on the idea that the Universe was comprised ultimately of software. We should not think of ultimate reality as particles and forces, according to Fredkin, but rather as bits of data modified according to computation rules.

Fredkin is quoted by Robert Wright in the 1980s as saying "There are three great philosophical questions. What is life? What is consciousness and thinking and memory and all that? And how does the Universe work? The informational viewpoint encompasses all three. . . . What I'm saying is that at the most basic level of complexity an information process runs what we think of as physics. At the much higher level of complexity, life, DNA - you know, the biochemical functions - are controlled by a digital information process. Then, at another level, our thought processes are basically information processing. . . . I find the supporting evidence for my beliefs in ten thousand different places, and to me it's just totally overwhelming. It's like there's an animal I want to find. I've found his footprints. I've found his droppings. I've found the half-chewed food. I find pieces of his fur, and so on. In every case it fits one kind of animal, and it's not like any animal anyone's ever seen. People say, where is this animal? I say, Well he was here, he's about this big, this that, and the other. And I know a thousand things about him. I don't have him in hand, but I know he's there. . . . What I see is so compelling that it can't be a creature of my imagination."[13]

In commenting on Fredkin's theory of digital physics, Robert Wright writes, "Fredkin . . . is talking about an interesting characteristic of some computer programs, including many cellular automata: there is no shortcut to finding out what they will lead to. This, indeed, is a basic difference between the "analytical" approach associated with traditional mathematics, including differential equations, and the "computational" approach associated with algorithms. You can predict a future state of a system susceptible to the analytic approach without figuring out what states it will occupy between now and then, but in the case of many cellular automata, you must go through all the intermediate states to find out what the end will be like: there is no way to know the future except to watch it unfold. . . There is no way to know the answer to some question any faster than what's going on. . . . Fredkin believes that the Universe is very literally a computer and that it is being used by someone, or something, to solve a problem. It sounds like a good-news / bad-news joke: the good news is that our lives have purpose; the bad news is that their purpose is to help some remote hacker estimate pi to nine jillion decimal places."[14]

Fredkin went on to show that although energy is needed for information storage and retrieval, we can arbitrarily reduce the energy required to perform any particular example of information

processing, and there is no lower limit to the amount of energy required[15]. This result made plausible the view that information rather than matter and energy should be regarded as the more fundamental reality.

I discussed Weiner's and Fredkin's view of information as the fundamental building block for physics and other levels of reality in my 1990 book The Age of Intelligent Machines[16].

The complexity of casting all of physics in terms of computational transformations proved to be an immensely challenging project, but Fredkin has continued his efforts.[17] Wolfram has devoted a considerable portion of his efforts over the past decade to this notion, apparently with only limited communication with some of the others in the physics community who are also pursuing the idea.

Wolfram's stated goal "is not to present a specific ultimate model for physics,"[18] but in his "Note for Physicists,"[19] which essentially equates to a grand challenge, Wolfram describes the "features that [he] believe[s] such a model will have."

In The Age of Intelligent Machines, I discuss "the question of whether the ultimate nature of reality is analog or digital," and point out that "as we delve deeper and deeper into both natural and artificial processes, we find the nature of the process often alternates between analog and digital representations of information."[20] As an illustration, I noted how the phenomenon of sound flips back and forth between digital and analog representations. In our brains, music is represented as the digital firing of neurons in the cochlear representing different frequency bands. In the air and in the wires leading to loudspeakers, it is an analog phenomenon. The representation of sound on a music compact disk is digital, which is interpreted by digital circuits. But the digital circuits consist of thresholded transistors, which are analog amplifiers. As amplifiers, the transistors manipulate individual electrons, which can be counted and are, therefore, digital, but at a deeper level are subject to analog quantum field equations.[21] At a yet deeper level, Fredkin, and now Wolfram, are theorizing a digital (i.e., computational) basis to these continuous equations. It should be further noted that if someone actually does succeed in establishing such a digital theory of physics, we would then be tempted to examine what sorts of deeper mechanisms are actually implementing the computations and links of the cellular automata. Perhaps, underlying the cellular automata that run the Universe are yet more basic analog phenomena, which, like transistors, are subject to thresholds that enable them to perform digital transactions.

Thus establishing a digital basis for physics will not settle the philosophical debate as to whether reality is ultimately digital or analog. Nonetheless, establishing a viable computational model of physics would be a major accomplishment. So how likely is this?

We can easily establish an existence proof that a digital model of physics is feasible, in that continuous equations can always be expressed to any desired level of accuracy in the form of discrete transformations on discrete changes in value. That is, after all, the basis for the fundamental theorem of calculus[22]. However, expressing continuous formulas in this way is an inherent complication and would violate Einstein's dictum to express things "as simply as possible, but no simpler." So the real question is whether we can express the basic relationships

that we are aware of in more elegant terms, using cellular-automata algorithms. One test of a new theory of physics is whether it is capable of making verifiable predictions. In at least one important way that might be a difficult challenge for a cellular automata-based theory because lack of predictability is one of the fundamental features of cellular automata.

Wolfram starts by describing the Universe as a large network of nodes. The nodes do not exist in "space," but rather space, as we perceive it, is an illusion created by the smooth transition of phenomena through the network of nodes. One can easily imagine building such a network to represent "naïve" (i.e., Newtonian) physics by simply building a three-dimensional network to any desired degree of granularity. Phenomena such as "particles" and "waves" that appear to move through space would be represented by "cellular gliders," which are patterns that are advanced through the network for each cycle of computation. Fans of the game of "Life" (a popular game based on cellular automata) will recognize the common phenomenon of gliders, and the diversity of patterns that can move smoothly through a cellular automaton network. The speed of light, then, is the result of the clock speed of the celestial computer since gliders can only advance one cell per cycle.

Einstein's General Relativity, which describes gravity as perturbations in space itself, as if our three-dimensional world were curved in some unseen fourth dimension, is also straightforward to represent in this scheme. We can imagine a four-dimensional network and represent apparent curvatures in space in the same way that one represents normal curvatures in three-dimensional space. Alternatively, the network can become denser in certain regions to represent the equivalent of such curvature.

A cellular-automata conception proves useful in explaining the apparent increase in entropy (disorder) that is implied by the second law of thermodynamics. We have to assume that the cellular-automata rule underlying the Universe is a Class 4 rule (otherwise the Universe would be a dull place indeed). Wolfram's primary observation that a Class 4 cellular automaton quickly produces apparent randomness (despite its determinate process) is consistent with the tendency towards randomness that we see in Brownian motion, and that is implied by the second law.

Special relativity is more difficult. There is an easy mapping from the Newtonian model to the cellular network. But the Newtonian model breaks down in special relativity. In the Newtonian world, if a train is going 80 miles per hour, and I drive behind it on a nearby road at 60 miles per hour, the train will appear to pull away from me at a speed of 20 miles per hour. But in the world of special relativity, if I leave Earth at a speed of three-quarters of the speed of light, light will still appear to me to move away from me at the full speed of light. In accordance with this apparently paradoxical perspective, both the size and subjective passage of time for two observers will vary depending on their relative speed. Thus our fixed mapping of space and nodes becomes considerably more complex. Essentially each observer needs his own network. However, in considering special relativity, we can essentially apply the same conversion to our "Newtonian" network as we do to Newtonian space. However, it is not clear that we are achieving greater simplicity in representing special relativity in this way.

A cellular node representation of reality may have its greatest benefit in understanding some aspects of the phenomenon of quantum mechanics. It could provide an explanation for the

apparent randomness that we find in quantum phenomena. Consider, for example, the sudden and apparently random creation of particle-antiparticle pairs. The randomness could be the same sort of randomness that we see in Class 4 cellular automata. Although predetermined, the behavior of Class 4 automata cannot be anticipated (other than by running the cellular automata) and is effectively random.

This is not a new view, and is equivalent to the "hidden variables" formulation of quantum mechanics, which states that there are some variables that we cannot otherwise access that control what appears to be random behavior that we can observe. The hidden variables conception of quantum mechanics is not inconsistent with the formulas for quantum mechanics. It is possible, but is not popular, however, with quantum physicists because it requires a large number of assumptions to work out in a very particular way. However, I do not view this as a good argument against it. The existence of our Universe is itself very unlikely and requires many assumptions to all work out in a very precise way. Yet here we are.

A bigger question is how could a hidden-variables theory be tested? If based on cellular automata-like processes, the hidden variables would be inherently unpredictable, even if deterministic. We would have to find some other way to "unhide" the hidden variables.

Wolfram's network conception of the Universe provides a potential perspective on the phenomenon of quantum entanglement and the collapse of the wave function. The collapse of the wave function, which renders apparently ambiguous properties of a particle (e.g., its location) retroactively determined, can be viewed from the cellular network perspective as the interaction of the observed phenomenon with the observer itself. As observers, we are not outside the network, but exist inside it. We know from cellular mechanics that two entities cannot interact without both being changed, which suggests a basis for wave function collapse.

Wolfram writes that "If the Universe is a network, then it can in a sense easily contain threads that continue to connect particles even when the particles get far apart in terms of ordinary space." This could provide an explanation for recent dramatic experiments showing nonlocality of action in which two "quantum entangled" particles appear to continue to act in concert with one another even though separated by large distances. Einstein called this "spooky action at a distance" and rejected it, although recent experiments appear to confirm it.

Some phenomena fit more neatly into this cellular-automata network conception than others. Some of the suggestions appear elegant, but as Wolfram's "Note for Physicists" makes clear, the task of translating all of physics into a consistent cellular automata-based system is daunting indeed.

Extending his discussion to philosophy, Wolfram "explains" the apparent phenomenon of free will as decisions that are determined but unpredictable. Since there is no way to predict the outcome of a cellular process without actually running the process, and since no simulator could possibly run faster than the Universe itself, there is, therefore, no way to reliably predict human decisions. So even though our decisions are determined, there is no way to predetermine what these decisions will be. However, this is not a fully satisfactory examination of the concept. This observation concerning the lack of predictability can be made for the outcome of most physical

processes, e.g., where a piece of dust will fall onto the ground. This view thereby equates human free will with the random descent of a piece of dust. Indeed, that appears to be Wolfram's view when he states that the process in the human brain is "computationally equivalent" to those taking place in processes such as fluid turbulence.

Although I will not attempt a full discussion of this issue here, it should be noted that it is difficult to explore concepts such as free will and consciousness in a strictly scientific context because these are inherently first-person subjective phenomena, whereas science is inherently a third person objective enterprise. There is no such thing as the first person in science, so inevitably concepts such as free will and consciousness end up being meaningless. We can either view these first person concepts as mere illusions, as many scientists do, or we can view them as the appropriate province of philosophy, which seeks to expand beyond the objective framework of science.

There is a philosophical perspective to Wolfram's treatise that I do find powerful. My own philosophy is that of a "patternist," which one might consider appropriate for a pattern recognition scientist. In my view, the fundamental reality in the world is not stuff, but patterns.

If I ask the question, 'Who am I?' I could conclude that, perhaps I am this stuff here, i.e., the ordered and chaotic collection of molecules that comprise my body and brain.

However, the specific set of particles that comprise my body and brain are completely different from the atoms and molecules than comprised me only a short while (on the order of weeks) ago. We know that most of our cells are turned over in a matter of weeks. Even those that persist longer (e.g., neurons) nonetheless change their component molecules in a matter of weeks.

So I am a completely different set of stuff than I was a month ago. All that persists is the pattern of organization of that stuff. The pattern changes also, but slowly and in a continuum from my past self. From this perspective I am rather like the pattern that water makes in a stream as it rushes past the rocks in its path. The actual molecules (of water) change every millisecond, but the pattern persists for hours or even years.

It is patterns (e.g., people, ideas) that persist, and in my view constitute the foundation of what fundamentally exists. The view of the Universe as a cellular automaton provides the same perspective, i.e., that reality ultimately is a pattern of information. The information is not embedded as properties of some other substrate (as in the case of conventional computer memory) but rather information is the ultimate reality. What we perceive as matter and energy are simply abstractions, i.e., properties of patterns. As a further motivation for this perspective, it is useful to point out that, based on my research, the vast majority of processes underlying human intelligence are based on the recognition of patterns.

However, the intelligence of the patterns we experience in both the natural and human-created world is not primarily the result of Class 4 cellular automata processes, which create essentially random assemblages of lower level features. Some people have commented that they see ghostly faces and other higher order patterns in the many examples of Class 4 images that Wolfram provides, but this is an indication more of the intelligence of the observer than of the pattern

being observed. It is our human nature to anthropomorphize the patterns we encounter. This phenomenon has to do with the paradigm our brain uses to perform pattern recognition, which is a method of "hypothesize and test." Our brains hypothesize patterns from the images and sounds we encounter, followed by a testing of these hypotheses, e.g., is that fleeting image in the corner of my eye really a predator about to attack? Sometimes we experience an unverifiable hypothesis that is created by the inevitable accidental association of lower-level features.

Some of the phenomena in nature (e.g., clouds, coastlines) are explained by repetitive simple processes such as cellular automata and fractals, but intelligent patterns (e.g., the human brain) require an evolutionary process (or, alternatively the reverse-engineering of the results of such a process). Intelligence is the inspired product of evolution, and is also, in my view, the most powerful "force" in the world, ultimately transcending the powers of mindless natural forces.

In summary, Wolfram's sweeping and ambitious treatise paints a compelling but ultimately overstated and incomplete picture. Wolfram joins a growing community of voices that believe that patterns of information, rather than matter and energy, represent the more fundamental building blocks of reality. Wolfram has added to our knowledge of how patterns of information create the world we experience and I look forward to a period of collaboration between Wolfram and his colleagues so that we can build a more robust vision of the ubiquitous role of algorithms in the world.

The lack of predictability of Class 4 cellular automata underlies at least some of the apparent complexity of biological systems, and does represent one of the important biological paradigms that we can seek to emulate in our human-created technology. It does not explain all of biology. It remains at least possible, however, that such methods can explain all of physics. If Wolfram, or anyone else for that matter, succeeds in formulating physics in terms of cellular-automata operations and their patterns, then Wolfram's book will have earned its title. In any event, I believe the book to be an important work of ontology.

---

[1] Wolfram, A New Kind of Science, page 2.

[2] Ibid, page 849.

[3] Rule 110 states that a cell becomes white if its previous color and its two neighbors are all black or all white or if its previous color was white and the two neighbors are black and white respectively; otherwise the cell becomes black.

[4] Wolfram, A New Kind of Science, page 4.

[5] The genome has 6 billion bits, which is 800 million bytes, but there is enormous repetition, e.g., the sequence "ALU" which is repeated 300,000 times. Applying compression to the redundancy, the genome is approximately 23 million bytes compressed, of which about half specifies the brain's starting conditions. The additional complexity (in the mature brain) comes from the use of stochastic (i.e., random within constraints) processes used to initially wire specific areas of the brain, followed by years of self-organization in response to the brain's interaction with its environment.

[6] See my book The Age of Spiritual Machines, When Computers Exceed Human Intelligence (Viking, 1999), the section titled "Disdisorder" and "The Law of Increasing Entropy Versus the Growth of Order" on pages 30 - 33.

[7] A computer that can accept as input the definition of any other computer and then simulate that other computer. It does not address the speed of simulation, which might be slow in comparison to the computer being simulated.

[8] Wolfram, A New Kind of Science, page 383.

[9] The seven commands of a Turing Machine are: (i) Read Tape, (ii) Move Tape Left, (iii) Move Tape Right, (iv) Write 0 on the Tape, (v) Write 1 on the Tape, (vi) Jump to another command, and (vii) Halt.

[10] As Wolfram points out, the previous simplest Universal Turing machine, presented in 1962, required 7 states and 4 colors. See Wolfram, A New Kind of Science, pages 706 - 710.

[11] The "nor" gate transforms two inputs into one output. The output of "nor" is true if an only if neither A *nor* B are true.

[12] See my book The Age of Intelligent Machines, section titled "A *nor* B: The Basis of Intelligence?," pages 152 - 157.

[13] Edward Fredkin, as quoted in Did the Universe Just Happen by Robert Wright.

[14] Ibid.

[15] Many of Fredkin's results come from studying his own model of computation, which explicitly reflects a number of fundamental principles of physics. See the classic Edward Fredkin and Tommaso Toffoli, "Conservative Logic," *International Journal of Theoretical Physics* 21, numbers 3-4 (1982). Also, a set of concerns about the physics of computation analytically similar to those of Fredkin's may be found in Norman Margolus, "Physics and Computation," Ph.D. thesis, MIT.

[16] See The Age of Intelligent Machines, section titled "Cybernetics: A new weltanschauung," pages 189 - 198.

[17] See the web site: http://www.digitalphilosophy.org, including Ed Fredkin's essay "Introduction to Digital Philosophy." Also, the National Science Foundation sponsored a workshop during the summer of 2001 titled "The Digital Perspective," which covered some of the ideas discussed in Wolfram's book. The workshop included Ed Fredkin Norman Margolus, Tom Toffoli, Charles Bennett, David Finkelstein, Jerry Sussman, Tom Knight, and Physics Nobel Laureate Gerard 't Hooft. The workshop proceedings will be published soon, with Tom Toffoli as editor.

[18] Stephen Wolfram, A New Kind of Science, page 1,043.

[19] Ibid, pages 1,043 - 1,065.

[20] The Age of Intelligent Machines, pages 192 - 198.

[21] Ibid.

[22] The fundamental theorem of calculus establishes that differentiation and integration are inverse operations.

What Have We Learned a Year After NASDAQ Hit 5,000?

Ray Kurzweil

http://www.kurzweilai.net/articles/art0379.html

The current recession reflects failure to develop realistic models of the pace at which new information-based technologies emerge and the overall acceleration of the flow of information. But in the longer-range view, recessions and recoveries reflect a relatively minor variability compared to the far more important trend of the underlying exponential growth of the economy.

Published on KurzweilAI.net January 21, 2002.

Although the Internet revolution is real and continues (e.g., continued exponential growth of e-commerce, the number of web hosts, the volume of Internet data, and many other measures of the power of the Internet), this does not change a fundamental requirement for business success: vertical market expertise. Most companies use the telephone, but we don't define them as telephone-centric companies. Your local dry cleaner is likely to have a web site today, so the web has become about as ubiquitous as the phone, but we still want a cleaner that knows something about cleaning clothes.

The real Internet revolution has been the adoption of decentralized Internet-based communication by traditional companies to redefine their internal work flow processes and to communicate up and down the supply chain including end users. However, the proper definition of an Internet company is one that makes distinctive use of the power of the network. A company like eBay, for example, would not be possible in the brick and mortar world and makes unique use of its ability to match buyers and sellers.

We also learned that although a new technology may ultimately be destined to profoundly affect our civilization, there are nonetheless well-defined limits at specific points in time to its varied requirements. On the order of a trillion dollars of lost market capitalization in telecommunications resulted from absurd over investment in some aspects of the technology (e.g., the extreme glut of fiber) before other enabling technologies (e.g., the "last mile" of user connectivity) were ready.

Because of improvements in communication between buyers and sellers, this recession is not about excessive inventory. It resulted instead from a failure to develop realistic models of the pace at which new information-based technologies emerge.

It is also the case that the pendulum is swinging more quickly now, which reflects the overall acceleration of the flow of information. We went from almost anything goes a year ago to almost nothing goes four months ago to signs today of a renewed willingness to invest in new ideas by the angel, venture capital and IPO communities.

An important phenomenon I've noted from the recessions of the twentieth century (including the Great Depression) is that the recessions and recoveries reflect a relatively minor variability

compared to the far more important trend of the underlying exponential growth of the economy. It is interesting to note that as each recession ended, the economy ended up exactly where it would have been (in terms of the underlying exponential growth) had the recession never occurred in the first place, as one can see in the following chart.

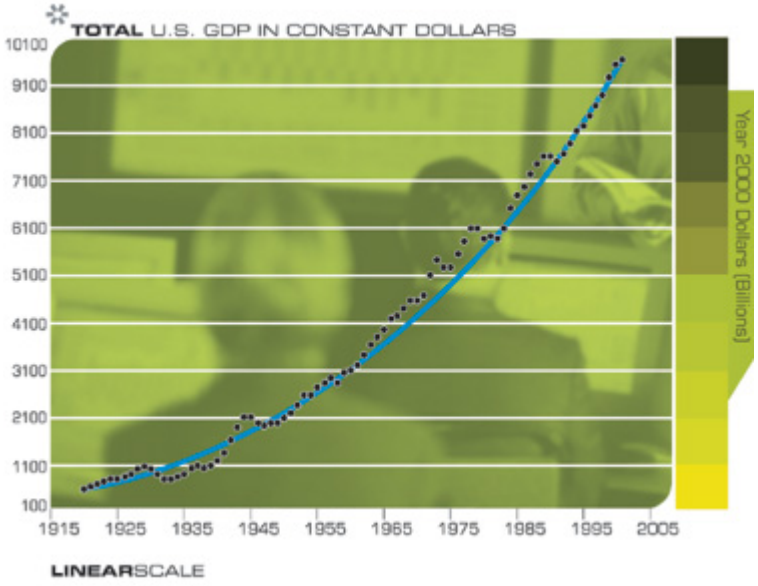


Chart Graphics by Brett Rampata/Digital Organism

Remarks on Accepting the American Composers Orchestra Award

Ray Kurzweil

http://www.kurzweilai.net/articles/art0356.html

The Second Annual American Composers Orchestra Award for the Advancement of New Music in America was presented on November 13 to Ray Kurzweil by American Composers Orchestra. Kurzweil reflects on creativity and the jump from the blackboard to changing peoples' lives.

Published on KurzweilAI.net November 14, 2001. Originally presented November 13, 2001. See related news item on KurzweilAI.net.

Music is the most universal form of expression known to human civilization, more so than other art forms such as dance, painting, and literature. Every known culture that has been discovered has expressed itself through music. What we express in music represents our most universal ideas, themes of life and death, of our connection to each other and to our spiritual origins.

While music has its roots in our primal history, it is also the case that music has always used the most advanced technologies available, from the cabinet making crafts of the eighteenth century, the metalworking industries of the nineteenth century, the analog electronics of the 1960s to the digital signal processing chips of the 1990s and early twenty-first century.

Music is both ancient and modern, and embraces both our folk traditions and our cutting-edge science. Music looks both backward and forward, and thereby embodies Winston Churchill's maxim that, "the further backward you look, the further forward you can see." Churchill's insight is a fitting citation for the American Composers Orchestra.

The ACO has dedicated itself to giving voice to innovative composers who are pioneering modern ways to apply all of our musical traditions and methods. Coming from a musical family and an upbringing that valued diverse musical idioms, it is a special honor for me to accept this unique and wonderful award.

The exciting thing for me as an inventor is that magical leap from dry formulas on a blackboard to actual transformations in people's lives. That's the delight of inventing. What my colleagues and I had tried to accomplish in the area of musical technology was to provide a technological bridge between the old world of acoustic instruments, and the new world of artistic control provided by synthesizers, sequencers, sound processors, and controllers.

Today we've broken the link between the physics of creating signals and the playing techniques required to generate sound. A musician today can apply any form of playing skill to produce any timbres, can create music in non-real-time, and can jam along with the focused intelligence of cybernetic musicians.

The pace of change is growing exponentially, and we can be sure that the means of creating music will accelerate as well. Of course, it takes more than technology to create music. Music will remain the expression of human ideas and emotions through the medium of sound. And we can be confident that the American Composers Orchestra will remain an inspiring center of innovation and excellence for musical creativity.

Thank you once again for this marvelous honor.

Foreword to The Eternal E-Customer

Ray Kurzweil

(book by Bryon Bergeron)

http://www.kurzweilai.net/articles/art0228.html

How have advances in electronic communications changed power relationships? The toppling of a government provides one not-so-subtle example. Ray Kurzweil talks about those advances in this forward to The Eternal E-Customer, a book that looks at the principles companies must adopt to meet the needs and desires of this new kind of customer.

Published on KurzweilAI.net July 6, 2001. Book originally published by McGraw Hill October 27, 2000.

The advent of worldwide decentralized communication epitomized by the Internet and cell phones has been a pervasive democratizing force. It was not Yeltsin standing on a tank that overturned the 1991 coup against Gorbachev, but rather the clandestine network of fax machines and early forms of e-mail that broke decades of totalitarian control of information. The movement toward democracy and capitalism and the attendant economic growth that has characterized the 1990s have all been fueled by the accelerating force of these person-to-person communication technologies.

The impact of distributed and intelligent communications has been felt, perhaps most intensely in the world of business. Despite dramatic mood swings on Wall Street, the seemingly extraordinary values often ascribed to so-called "e-companies" reflects a genuine perception: the business models that have sustained businesses for decades are in the early phases of a radical transformation. New models based on direct personalized communication with the customer will transform every industry, resulting in massive disintermediation of the middle layers of distribution that have traditionally separated the customer from the ultimate source of products and services.

The underlying technologies are all accelerating. It's not just computation that is growing exponentially, but also communication, networks, biological sciences (e.g., DNA sequencing), brain scanning, miniaturization (we are currently shrinking technology at a rate of 5.6 per linear dimension per decade), the accumulation of knowledge, and even the rate of paradigm shift itself. And the underlying technologies are becoming ever more intelligent, subtle, emotionally aware, that is, more human.

Expanding access to knowledge is changing power relationships. Patients increasingly approach visits to their physician armed with a sophisticated understanding of their medical condition and their options. Consumers of virtually everything from toasters, cars, and homes to banking and insurance are now using automated software agents ("bots") to quickly identify the right choices with the optimal features and prices.

The wishes and desires of the customer, often unknown even to herself, are rapidly becoming the driving force in business relationships. The well connected clothes shopper, for example, is not going to be satisfied for much longer with settling for whatever items happen to be left hanging on the rack of her local store. Instead, she will select just the right materials and styles by viewing how many possible combinations look on an image of her own body (based on a detailed three-dimensional body scan), and then having her choices custom manufactured.

The current disadvantages of web-based commerce (e.g., limitations in the ability to directly interact with products and the frustrations of interacting with inflexible menus and forms instead of human personnel) will gradually dissolve as the trends move robustly in favor of the electronic world. By the end of this decade, computers will disappear as distinct physical objects. Displays will be written directly onto our retinas by devices in our eyeglasses and contact lenses. In addition to virtual high resolution displays, these intimate displays will provide full immersion visual virtual reality. We will have ubiquitous very high bandwidth wireless connection to the Internet at all times. "Going to a web site" will mean entering a virtual reality environment - at least for the visual and auditory sense - where we can directly interact with products and people, both real and simulated. Although the simulated people will not be up to human standards, not by 2009, they will be quite satisfactory as sales agents, reservation clerks, and research assistants. The electronics for all of this will be so small that it will be invisibly embedded in our glasses and clothing. Haptic (i.e., tactile) interfaces will enable us to touch products and people. It is difficult to identify any lasting advantage of the old brick and mortar world that will not ultimately be overcome by the rich interactive interfaces that are soon to come.

If we go further out — to, say 2029, as a result of continuing trends in miniaturization, computation, and communication, we will have billions of nanobots - intelligent robots the same of blood cells or smaller - traveling through the capillaries of our brain communicating directly with our biological neurons. By taking up positions next to every nerve fiber coming from all of our senses, the nanobots will provide full immersion virtual reality involving all five of the senses. So we will enter virtual reality environments (via the web, of course) of our choice, interact with a panoply of intelligent products and services, and meet people, both real and virtual, only now the difference won't be so clear.

In his brilliant and entertaining book, Bryan Bergeron has provided a comprehensive and insightful roadmap to this e-revolution now in its infancy. Dr. Bergeron describes this era not as a single transformation, but as an ongoing churning that will continually uproot and exchange one set of business models for another. What is needed, Bryan tells us, is the right set of principles that can enable businesses to flourish through times of ever accelerating change. He discerningly bases these principles on the loyalty of the increasingly empowered customer. My advice would be to invest in any company that can successfully adopt Bryan Bergeron's principles of meeting the needs and desires of "the eternal e-customer."

The Eternal E-Customer: How Emotionally Intelligent Interfaces Can Create Long-Lasting Customer Relationship

Response to Fortune Editor's Invitational

Ray Kurzweil

http://www.kurzweilai.net/articles/art0222.html

Ray Kurzweil was invited to participate in the 2001 Fortune Magazine conference in Aspen, Colorado, which featured luminaries and leaders from the worlds of technology, entertainment and commerce. Here are his responses to questions addressed at the conference.

Published on KurzweilAI.net July 11, 2001. Originally written July 2001 to be presented August 2001.

Once upon a time we committed ourselves to putting a man on the moon. What kind of similar commitment should we make now? What's the "moon shot" of the 21st century?

Create technology that combines the strengths of human and machine intelligence and implement it in both our machines and in ourselves.

We biological humans excel in our powers of pattern recognition as well as our emotional intelligence. Our ability to recognize and respond appropriately to emotion reflects the most complex and subtle thing we do. Machine intelligence also has salient advantages, for example the ability to instantly share knowledge, speed (electronics is already 10 million times faster than our interneuronal connections) and capacity.

Today, our human intelligence is restricted to a mere 100 trillion connections in a human brain. Although the capacity of our computers today is still millions of times less powerful than the human brain, the basic architecture of our biological nervous system is fixed whereas the price-performance of computation is expanding exponentially. The cross-over point is only decades away.

Moreover, the software of human-level intelligence is not hidden from us. We are also making exponential gains in scanning, modeling, and reverse engineering the human brain, which will be the ultimate source for the "methods" of intelligence.

We will ultimately have the opportunity to combine the rich, diverse, and flexible powers of human intelligence with the knowledge sharing, speed, and capacity of machine intelligence. What form will this take? The answer is many different forms. One mode will be fully nonbiological entities with human-like qualities. The more interesting prospect will be expanding our own thinking through intimate connection with machine intelligence.

This undertaking will be the result of the ongoing exponential growth of computation and communication, the continuing shrinking of technology, as well as our accelerating understanding of the human brain and neural system. This should not be a NASA-style government project, but rather should reflect the ongoing interplay of private enterprise with a panoply of academic and government research institutions.

As the world heads down its current path, what should we fear most?

Technology amplifies both our creative and destructive natures. Our lives are immeasurably better off today than 100 years ago, but the twentieth century has also witnessed great amplification of our means for destruction.

Most powerful and potentially pernicious is self-replication. We already have self-replication in the medium of nuclear processes, and we are on the threshold of widely available means for creating bioengineered pathogens. In a couple of decades, we'll also have the ability to create self-replicating nonbiological entities in which key features are measured in nanometers. Following that will be the emergence of nonbiological intelligence smart enough to invent its own next generation suggesting a run away phenomenon not clearly under the control of biological humanity. Twenty-first century technologies will be billions of times more powerful than those of the twentieth century, and there are a myriad of downside scenarios that we can already envisage.

Calls for relinquishing potentially dangerous technologies such as nanotechnology are not the answer. For one thing, nanotechnology is not a unified field, but rather the inevitable end result of a broad trend toward miniaturization that pervades most areas of technology. We could scarcely stop the emergence of nanotechnology without "relinquishing" virtually all technology development. Moreover, the dangerous technologies represent the same knowledge as the beneficial ones. For example, the same biotechnologies that will save millions of lives from cancer and other diseases in the years ahead is precisely the same know-how that can potentially empower a terrorist to create a new pathogen.

Although I believe the risks are real, I believe that maintaining a free and open society is our best route to developing effective countermeasures. Serious attempts to relinquish broad areas of knowledge will only drive them underground where the less responsible practitioners (i.e., the terrorists) will have all the expertise.

What issue or issues will most define our future?

- **Who or what is human.** The exponential growth of information-based technologies (computation, communications, varied biotechnologies, human brain scanning and reverse engineering) combined with the exponential shrinking of the size of technology will blur the line between human and machine.
- **How can we avoid grave new dangers of technology while reaping profound benefits.** Most observers use linear extrapolation for their estimates of future time-frames, but this ignores the exponential nature of progress. We're currently doubling the paradigm shift rate every decade. This will raise the issue of how can we reliably anticipate the impact of technology to emphasize the promise while we avoid the peril.

How do we reinvent social institutions when people rarely die. Human life span is also growing exponentially through multiple biotechnology revolutions (Genomics, Proteomics, therapeutic cloning, rational drug design, and others), to be followed a couple of decades hence by human body and brain augmentation through nanotechnology. Within ten years, we'll be

adding more than a year every year to human life expectancy. Many issues will be raised as all of our human traditions need to be rethought.

The Singularity

“The Singularity” is a phrase borrowed from the astrophysics of black holes. The phrase has varied meanings; as used by Vernor Vinge and Ray Kurzweil, it refers to the idea that accelerating technology will lead to superhuman machine intelligence that will soon exceed human intelligence, probably by the year 2030.

KurzweilAI.net News of 2002

Ray Kurzweil, Amara Angelica

http://www.kurzweilai.net/articles/art0550.html

In its second year of operation, 2002, KurzweilAI.net continued to chronicle the most notable news stories on accelerating intelligence. Ray Kurzweil offers here his overview of the dramatic progress that the past year has brought.

Published on KurzweilAI.net Feb. 5, 2003

The capital markets (venture and angel financing, IPOs, mergers and acquisitions, bank debt) were in a deep freeze during 2002, particularly for high-tech ventures. Technology market capitalizations continued their downward trend from the previous year, and many ventures ended their operations for lack of funding.

So one might assume that this high tech recession might have slowed the pace of progress. A frequent question I (Ray) receive at my lectures is what impact the high-tech meltdown has had on my law of accelerating returns. Surely, many people point out, the acceleration of computation and other technologies must have been negatively affected.

The reality is that the law of accelerating returns is alive and well. We see no impact of either the boom times or the bust on the ongoing and unperturbed acceleration of the power of technology. The doubling of price-performance in a wide range of information technologies, including computation, magnetic and semiconductor memories, wireless and wired communication, miniaturization of electronics and mechanics, genomic sequencing, neural scanning, brain reverse engineering, and many others, has continued unabated.

The pace of innovation itself has also been undeterred. The unavailability of investment has served primarily to weed out poorly grounded ideas and projects. The spirit of innovation is so deeply embedded in our society and culture that creativity has not only continued to flourish, but even the number of fields and types of applications continue to multiply.

All of this is eminently apparent from the depth and diversity of last year's torrent of groundbreaking news as chronicled here on KurzweilAI.net. We set a high bar for news stories, and nonetheless posted 823 of them—more than two each day. We've selected just over half of these below to document the key breakthroughs in the continued exponential growth of these increasingly diverse information-based technologies; our deepening understanding of the information basis of biological processes; and the early contributions of nanotechnology.

A frequent challenge that I (Ray) receive is that while it is apparent that hardware technologies are growing exponentially, the same is not true for software. To this end, we hear an endless litany of frustrations with poorly designed software, and the complaint that software is no more intelligent today than it was years ago. My own view is that the "doubling time" for software productivity is indeed slower than that for hardware (I estimate it to be around five years versus

one year for hardware), but we are indeed benefiting from all the investments in new languages, class libraries, and development tools.

Smart software and consumer robots

Consider one dramatic example that we saw in 2002. It is often said that game-playing programs, such as the chess machines Deep Blue and Deep Fritz, rely entirely on brute force expansion of the move-countermove tree. But there is an important aspect of these programs that requires qualitative intelligence, namely the evaluation of the "terminal leaves" of this tree. It would not be an intelligent use of computer time to endlessly expand every branch of the tree. For example, if one side was down by a queen and a rook at a particular node in a branch, there would be little point in considering that line of play further. So the classical "minimax" game-playing strategy does require an important judgment at each node of the tree: should we abandon or continue the expansion?

Deep Blue, which defeated Gary Kasparov, the human world champion, in 1987, used a set of several hundred finely tuned parameters to make this delicate decision. Its hardware consisted of specialized chess circuits that were able to analyze 200 million board positions per second. Deep Fritz, which competed during 2002, is simply a software program running on eight conventional PC-class processors, and thus is only able to consider about three million board positions per second.

Despite this, it ranked about equal to Deep Blue and fought Vladimir Kramnik, the contemporary human world champion, to a draw. This impressive performance, despite using a small fraction of the computation used by Deep Blue, is entirely due to important qualitative improvements in the intelligence of its pruning software.

AI programs played a key role in many practical applications. Every time you send an email or place a cell phone call, you are calling upon increasingly sophisticated software programs to route your communication. Every time you make a credit card purchase, AI-based algorithms are analyzing the patterns of your purchases to look for fraud. AI programs have also been deployed on the emerging homeland security front, featuring the ability to detect patterns that identify potential terrorist behavior, bioterrorism outbreaks, and terrorist or criminal movements.

In *The Age of Intelligent Machines*, which I wrote in the late 1980s, I predicted that warfare would be transformed from the variables that had dominated strategy from ancient times — geography, offensive firepower, and defensive fortifications —to the sophistication of intelligent software and communications. We were headed towards an era in which combat would increasingly be conducted by intelligent machines, with humans (at least those on the winning side) becoming increasingly removed from the scene of conflict.

The Persian Gulf War of 1991 saw the early use of intelligent weapons, although only about five percent of our munitions were "smart." In this past year, the majority of our weapons were "smart" weapons during the Afghanistan conflict, and estimates are that 95 percent of our munitions will be intelligent weapons for the Iraqi conflict of 2003 (assuming that President Hussein does not back down in the face of our smarter electronics and software).

In 2002, we saw the Pentagon turn to terrestrial robots and airborne drones for spying, detecting land mines, and combat in Afghanistan. Looking ahead, researchers are actively developing smarter and smaller weapons, for example, flying robots modeled on insects and birds.

Robotics made notable advances on the civilian side as well, including research on sociable robots, ones that can sense human emotions, highly mobile robots modeled on cockroaches, a mobile robot that can learn in real time by matching images with its memory, and one controlled by a hybrid rat-silicon brain. Notably one robot taught itself the principles of flying by trial and error in just three hours.

Consumer robots also made news, including a housecleaning robot, a realistic cat toy, a walking-talking Honda car salesman, and—shades of the movie *A.I.*—a child robot that can interact with its "carers," expressing emotions.

Robots also found another popular consumer role as "robodocs" in elder care facilities and homes, in the form of cuddly bears and AIBO dogs to offer social stimulation and devices to monitor patients' medical condition and behavior and give directions to Alzheimers patients when lost. Remote-control robots were also used by doctors for precision heart surgery.

There were also improvements in speech recognition, speech synthesis, and natural language processing.

A new category of business ventures focuses on developing virtual personalities that will handle routine transactions over the phone, such as making reservations and conducting purchases and performing information queries. Many of these systems were rolled out during 2002.

For example, you can call British Airways and talk to their automated attendant about anything you want, as long as it has to do with booking a flight with British Airways. Other developments included optimized dictation along with voice-controlled email and Web surfing, a talking book for the blind, full-text search of audio recordings, and a DARPA program to develop a handheld computer capable of speech recognition and translation between 13 languages in four subject areas.

Several computing technology breakthroughs were announced, including a three-centimeter disc that stores four gigabytes of data or video, a magnetic film-based hard drive that stores 200 gigabytes per square inch, and a 12-cm, CD-size disc that stores one terabyte of data.

Have supercomputers already achieved human brain capacity?

Supercomputing growth continued Moore's Law in 2002, with power for the same price doubling every 15 months. The world's fastest computers are fast approaching the computational capacity of the human brain. There are various estimates of this capacity, with the one I have offered being among the conservatively high estimates (10^{11} neurons times 10^{3} average fan-out times 10^{2} transactions per second = ~ 10^{16} transactions per second). Japan's Earth Simulator supercomputer became the world's fastest, with 35.86 teraflops (3 x 10^{13}) performance. Not to be outdone, IBM

announced plans to build the 100 teraflops (10^{14}) ASCI Purple and the 367 teraflops (3×10^{14}) Blue Gene/L.

Research on brain reverse engineering and neuromorphic modeling (for example, Lloyd Watt's work on emulating the auditory processing regions of the brain) has shown that neuromorphic modeling has the potential to reduce the computational requirement by at least 10^3, which means that we have already achieved human brain capacity, in the context of using neuromorphic models (with reduced computational requirements), and using supercomputers. Now all we need is the software of intelligence.

One of the keys to achieving the software of human intelligence is reverse-engineering the human brain. More powerful tools are emerging. For example, researchers at the University of Pennsylvania built a system that is able to noninvasively image up to 1,000 neurons in a layer of only 10 micrometers and at depths of up to 150 micrometers below the surface. Their goal is to achieve millisecond time scales of the activity of individual neurons in large clusters (1,000 or more) of neurons. This type of system will allow the development of detailed models of how neuron clusters learn new patterns of information.

With two million new Internet users per month and more than half of the U.S. population able to access the Web, pervasive computing is moving closer to becoming a reality, notably via wireless high-speed Wi-Fi (802.11) access to the Internet in public and private spaces. Simplifying use by consumers, AT&T, Intel and IBM announced a nationwide service integrating Wi-Fi with broadband cell-phone-based Internet access.

Blurring reality

The movie "Simone" brought the notion of a virtual reality actor (synthespian) to the public last year. This movie was of particular interest to me (Ray) because I am one of the few people in the world to have actually had the experience that the character Viktor Taransky (played by Al Pacino) has in the movie: transforming oneself in realtime into another person.

During the past year, researchers made strides in achieving several elements of this technology:

- virtual stunt artists that respond to the physics of the real world
- the first realistic videos of people saying things they never said
- techniques to allow a biomechanically realistic 3D model of a character to learn how to produce its own body motion
- a digital image sensor that was the first to match or surpass the photographic capabilities of 35-millimeter film
- software that converts standard video images into 3-D in real time
- a new technique for creating large, highly realistic holograms.

Augmented-reality systems were also developed to give surgeons critical data during operations, provide visual fly-throughs of a living tumor, achieve 3-D fractal computer modeling to construct vascular systems in artificial organs, speed up research into diseases by creating 3-D models of cells in a room similar to the Star Trek Holodeck, and offer stroke patients hand-impairment therapy.

Prototypes of innovative computer-display and video systems were announced, including flexible "electronic paper," an entire computer printed on glass, "smart displays" that wirelessly communicate with personal computers, a holographic video recorder, and picture-editing tools that can automatically trace outlines, seamlessly cover marks or blemishes, and fill in backgrounds when pieces of an image are removed.

With rising threats of terrorism and hacker attacks, innovative cybersecurity countermeasures were developed last year:

- real-time 3-D images for surveillance
- quantum encryption moved closer (keys cannot be intercepted without the sender and receiver knowing)
- computer-surveillance system to give U.S. counterterrorism officials access to personal information in government and commercial databases
- intrusion-detection software that mimics biological immune systems by learning to watch for unusual events
- RFID (Radio Frequency ID) tags for visitors to Saudi Arabia for logistics, crowd control, and security.

Beyond Moore's law

Intel chairman Andy Grove warned that as chips become increasingly dense, heat developed by current leakage will become a limiting factor to the growth described in Moore's law. Within a decade, he said, we'll need other solutions.

In the near term, 3-D chips may offer a solution. Several research labs (notably in Japan and at IBM) kicked off serious R&D efforts last year to develop 3-D chips. These vertically integrated devices promise lower prices while boosting power and speed. Leading the pack, Matrix Semiconductor plans to market its 3-D memory chip in the first half of 2003.

In the longer range, nanocomputing based on molecular electronic components became a leading candidate to replace conventional lithography. Silicon nanowires and carbon nanotubes are now the candidate nanoscale technologies that could begin to replace standard transistors in the decade after 2010. Research during the year included:

- 3-D nanotubes
- atomic-scale "peapods" made of buckyballs
- nanotubes that self-assemble into circuit elements
- boron crystalline nanowires ("nanowhiskers")
- a new laser-stamping technique that could produce ten-nanometers-wide features, allowing for 100 times more transistors on a chip
- atom-thin (.5 nm) layers of crystalline silicon called "quantum wells" that can exploit quantum properties on the atomic level to develop ultrafast transistors.
- superlattices—a series of silicon p-n junctions—in a single nanowire for creating highly integrated logic circuits, nanoscale LEDs, and photonic waveguides
- the world's smallest transistor, just nine nanometers in length, designed by IBM researchers

- a way to store 1,024 bits of information in 19 hydrogen atoms in a single liquid-crystal molecule.

Similar research advances in memory were realized, such as IBM's project to create nanotech-based data storage density of a trillion bits per square inch, using thousands of nano-sharp tips to punch ten-nanometer-wide indentations representing individual bits into a thin plastic rewriteable film.

Research also focused on quantum computing, which could work synergistically with nanocomputing to solve important new classes of problems that are impractical without the use of quantum computing's ambiguous "qu-bits." Research included a method of creating a reversible quantum phase transition in a Bose-Einstein condensate (a new state of matter), a crystal that traps light, quantum dots created and held together by genetically-engineered viruses, microelectronic "spintronics" devices that use the spin of the electron to store and compute data, and superconducting junctions.

Despite nascent efforts to ban nanotechnology research because of fears it might lead to nanowarfare and "grey goo" scenarios, revolutionary nanotech research moved ahead, including:
- longer-lasting batteries
- methods of fighting weapons of mass destruction by analyzing trace pathogens and chemicals
- self-mending and self-cleaning plastics, coatings and materials that resist friction and wear or shed dirt
- super-strong electrically-conducting threads

In addition, new federal legislation proposed in 2002 could result in $37 billion for research in nanotech, biotech, and other key new technologies. Researchers also developed breakthroughs in bionanotechnology—hybrid nanoscale devices based on biological molecules—including:
- viruses studded with molecules of gold and antibodies that could invade tumors and help assemble supercomputers
- protein-based nanoarrays for diagnosing infectious diseases and biological agents
- molecular motors using ATPase enzyme molecules attached to metallic substrates
- 50-nanometer spots of DNA that could create a gene-reading chip with 100,000 different diagnostic tests in an area the size of the tip of a needle in a few seconds
- bacteria to form microbial machines to repair wounds or build microscopic electrical circuits
- radio-controlled DNA that could act as electronic switches that allow scientists to turn genes on and off by remote control
- DNA to build nanorobots that could then build new molecules and computer circuits or fight infectious diseases.

Researchers also made breakthroughs in nanomedicine, including a smart membrane containing silica nanotubes capable of separating beneficial from useless or harmful forms of a cancer-fighting drug molecule, nanoparticles that cut tumors' supply lines, "tecto-dendrimers" for diseased-cell recognition, and Buckyballs with chemical groups attached for drug delivery.

Cyborgs, clones, and the cosmos

A number of new technologies were introduced in 2002 for creating cyborgs, dramatized by an experiment by "the world's first cyborg," Professor Kevin Warwick of the University of Reading, who implanted a microchip in his arm. Other developments included:

- "bionic" body replacement parts
- microelectronic retinal prostheses
- ID chip implants
- electroactive polymers to form "artificial muscles"
- a device that stimulates the visual cortex of the brain with video from a camera
- a powered exoskeleton for "supersoldiers"
- a method for shielding pacemakers against interference from MRI machines
- a touch display for the visually impaired
- implantable microchips for controlling robots with the mind

There were numerous biotech breakthroughs in 2002:

- Genome entrepreneur Craig Venter announced a service to map a person's entire genetic code and a plan to create a single-celled, partially man-made organism with enough genes to sustain life
- A prototype tool for half-hour DNA tests (rather than two weeks) from saliva
- Rat heads grafted onto the thighs of adult rats to investigate how the transplanted brain can develop and maintain function after prolonged total brain ischemia, which will help understand brain injury in newborn babies
- "Junk DNA" found to contain instructions essential for growth and survival
- Recently discovered "small RNA" molecules named by Science Magazine as the science breakthrough of the year. These operate many of the cell's controls and can shut down genes or alter their levels of expression.

Serious cloning research flourished in spite of the dubious announcement of the birth of the first human clone. Professor Ian Wilmut, who cloned Dolly the sheep, applied for a government license to work with human eggs in an experiment that would prepare the way for human cloning. The first cloned cat was successfully created. And five clone calves were born with 0.1 percent human DNA intended to produce C-1 Esterase Inhibitor to treat humans suffering from angioedema.

There was also significant progress in the related (also controversial) field of stem-cell research. Small RNA (mentioned above) may provide us with the key to achieve the holy grail of somatic cell engineering: directly transforming one type of cell into another. By manipulating the protein codes in the small RNA molecules that tell a cell what type of cell it is, we could create new cells, tissues, and organs directly from skin cells without the use of embryonic stem cells.

A major advantage of this approach is that the new tissues will have the patient's own DNA and thereby avoid autoimmune rejection. As an important step in that direction, during 2002, scientists found a way to transform skin cells into another type, including immune system cells and nerve cells, without using cloning or embryonic stem cells.

Researchers also:

- reversed symptoms of Parkinson's disease in rats using stem cells from mouse embryos
- isolated a stem cell from adult human bone marrow that can produce all the tissue types in the body
- grew functional kidneys using stem cells from cloned cow embryos
- discovered fetal stem cells and adult cells that can create neurons to repair a damaged brain

There was also important research progress in neuroscience in 2002, including:
- a "brain cap" to help assess astronauts' mental performance in orbit
- a system that noninvasively detects patterns of nerve connections inside the brains of living people
- electrodes attached to a single neuron in the motor cortex that allow for moving a cursor on a computer screen just by thinking about it
- a method of repairing brain damage in humans caused by stroke or brain tumors
- "brain pacemakers" for Parkinson's disease and other conditions
- transcranial magnetic stimulation to treat depression
- a chip patterned on the human eye that picks out the kinds of features and facial patterns that we use to recognize people and read their emotional state

The NSF also proposed a major research program to enhance human performance, such as developing broad-bandwidth interfaces directly between the human brain and machines.

New forms of energy were also developed, including a micro fuel cell that runs on methanol and provides much longer life than any other portable battery, the world's first commercially available cars running on hydrogen fuel cells, a new fuel cell that generates electricity from the glucose-oxygen reaction that occurs in human blood for powering medical sensors, and tiny batteries that could provide 50 years of power, drawing energy from radioactive isotopes.

There were also dramatic theoretical developments in cosmology: signals that appear to be transmitted at least four times faster than the speed of light (although this experimental result does not appear to allow the transmission of information at these speeds) and observations of black holes that suggest the speed of light is slowing.

Once the intelligence of our civilization spreads to other parts of the Universe, the maximum speed with which that influence can spread will become a critical consideration. There were hints this past year that the speed of light may not be an absolute barrier to reaching the far corners of the Universe in a reasonable period of time. Of particular interest were analyses showing the theoretical feasibility of quantum wormholes, which may offer short cuts to the rest of the cosmos.

Top KurzweilAI.net items in the following areas can be found at http://www.kurzweilai.net/articles/art0550.html:

3D Chips
Artificial Intelligence
Bionanotechnology
Biotechnology

Cloning and Stem Cell Research
Cosmology
Cybersecurity
Cyborgs
Displays
Energy
Moore's Law
Nanocomputing
Nanomedicine
Nanotechnology
Neuroscience
Pervasive Computing
Quantum Computing
Robodocs
Robotics
Speech Recognition and Synthesis
Super Computing
Virtual Reality

Singularity Math Trialogue

Ray Kurzweil, Vernor Vinge, and Hans Moravec

http://www.kurzweilai.net/articles/art0151.html

Hans Moravec, Vernor Vinge, and Ray Kurzweil discuss the mathematics of The Singularity, making various assumptions about growth of knowledge vs. computational power.

Published on KurzweilAI.net March 28, 2001

From: Hans Moravec

Date: February 15, 1999

To: Ray Kurzweil

Subject: Foldover in the accelerating exponential

Hi Ray,

Following up a thread touched on in your visit: Vernor Vinge noted (and wrote SF stories) in the 1960s that AI, by automating the process of innovation, would increasingly accelerate its own development, shortening the time constant of its own exponential.

If the doubling rate is proportional to the magnitude of the exponential, then the curve undergoes equal multiplicative steps in exponentially decreasing time intervals, and reaches infinite slope in a finite time.

Vinge calls this the "Technological Singularity," and notes that it blocks prediction, just as a mathematical singularity prevents extrapolation. There is probably a curve beyond the singularity, but it has a different character.

Here is a recent pointer:

http://www-personal.engin.umich.edu/~jxm/singlar.html

I looked at the prescription for the accelerating exponential, and noted the following:

From: Hans Moravec

Date: September 30, 1997

To: Vernor Vinge

Subject: A kind of Singularity

Hi Vernor,

A tidbit you've probably noticed before:

V = exp(V*t) has a singularity (at t = 1/e, V = e) in dV(t)/dt but not in V(t) itself. Instead, V folds back, making a double valued function that kind of suggests time travel!

You can see the curve most easily by parametrically plotting V against t = log(V)/V with the parameter V going from about 1 (or less) to 5.

Or take V only up to e to be neat.

Best—Hans

To: Hans Moravec

Subject: Re: A kind of Singularity

Date: October 1, 1997

From: Vernor Vinge

Hi Hans —

I hadn't noticed that. Neat and weird: "After a while the world disappeared, but only because it had already become very much nicer"!

Hope things are all going well. (I'm back to teaching after two years at writing. It is very interesting—and as hectic as I remember!)

Regards,

—Vernor

From: Hans Moravec

Date: Monday, February 15, 1999

To: Vernor Vinge, Ray Kurzweil

Subject: Singularity equation correction

So, I reconsidered the V = exp(V*t) formula that suggested a foldback in the progress curve, and decided it didn't make a lot of sense, in the way it jumbled time t and the rate V of computing.

More careful formulation makes for a less surprising conclusion.

Making maximal simplifying assumptions, and shifting and scaling all quantities to avoid constants:

Let W be "world knowledge", and assume that each additive increment in W results in a multiplicative improvement in miniaturization, and thus in computer memory and speed V. So:

$V = \exp(W)$

In the old days, assume an essentially constant number of humans worked unassisted at a steady pace to increase W at a steady rate:

$dW/dt = 1$

So $W = t$ and $V = \exp(t)$

which is a regular Moore's law.

Now, suppose instead W is created solely by computers, and increases at a rate proportional to computer speed. Then:

$dW/dt = V$ giving $dW/\exp(W) = dt$

This solves to $W = \log(-1/t)$ and $V = -1/t$

W and V rise very slowly when $t \ll 0$, might be mistaken for exponential around $t = -1$, and have a glorious singularity at $t = 0$.

From: Hans Moravec

Sent: March 10, 1999

To: Ray Kurzweil

Cc: Vernor Vinge

Subject: Response to Singularity Equations

Goody! This led to some new insights.

Ray said:

One of your assumptions is:

$V = \exp(W)$ (i.e., that computer power grows exponentially with world knowledge).

I don't think this is a well-founded assumption. Why would additive increments in W result in a multiplicative improvement in V? I agree that V grows with W, but it makes more sense to say that: if we double world knowledge, we double our understanding of how to build a computer, and, therefore, double the power of computation per unit cost.

The assumption that V = exp(W) is surely too optimistic. I was thinking in terms of independent innovations. For instance, one might be an algorithmic discovery (like log N sorting) that lets you get the same result with half the computation. Another might be a computer organization (like RISC) that lets you get twice the computation with the same number of gates. Another might be a circuit advance (like CMOS) that lets you get twice the gates in a given space. Others might be independent speed-increasing advances, like size-reducing copper interconnects and capacitance-reducing silicon-on-insulator channels. Each of those increments of knowledge more or less multiplies the effect of all of the others, and computation would grow exponentially in their number.

But, of course, a lot of new knowledge steps on the toes of other knowledge, by making it obsolete, or diluting its effect, so the simple independent model doesn't work in general. Also, simply searching through an increasing amount of knowledge may take increasing amounts of computation. I played with the V=exp(W) assumption to weaken it, and observed that the singularity remains if you assume processing increases more slowly, for instance V = exp(sqrt(W)) or exp($W^{1/4}$). Only when V = exp(log(W)) (i.e., = W) does the progress curve subside to an exponential.

Actually, the singularity appears somewhere in the I-would-have-expected tame region between and V = W and V = W^2 (!)

Unfortunately the transitional territory between the merely exponential V=W and the singularity-causing V=W^2 is analytically hard to deal with. I assume just before a singularity appears, you get non-computably rapid growth!

Your assumption:

N = $C4^{(C5*t)}$ (the number of computing devices is growing at its own exponential rate) is pretty arbitrary. Wouldn't it make more sense to have the number increase as some function of W? In the latter case, the number of computers could simply be factored into the V=f(W) equation (where my V means the total amount of computation in the world, equivalent to your N*V).

Suppose computing power per computer simply grows linearly with total world knowledge, but that the number of computers also grows the same way, so that the total amount of computational power in the world grows as the square of knowledge:

V = W*W

also dW/dt = V+1 as before

This solves to W = tan(t) and V = $\tan(t)^2$,

which has lots of singularities (I like the one at t = pi/2).

I also question the application of negative time, and in particular zero time (which provides for the singularity).

Oh, that's just a matter of labeling the axes, with the origin chosen in that particular place to avoid unsightly constants. You could shift it anywhere, if you replace the t in the formulas by (t-t0). But I like it where it is.

If there is a singularity, it's kind of natural to divide time into BS (the negative times before the singularity) and AS (the strange times afterwards). (I do worry a little that in some of my constant-free formulas, it is easy to find regions where W<0, though they can mostly be shifted away.)

—Hans

After the Singularity: A Talk with Ray Kurzweil

Ray Kurzweil

http://www.kurzweilai.net/articles/art0451.html

John Brockman, editor of Edge.org, recently interviewed Ray Kurzweil on the Singularity and its ramifications. According to Ray, "We are entering a new era. I call it 'the Singularity.' It's a merger between human intelligence and machine intelligence that is going to create something bigger than itself. It's the cutting edge of evolution on our planet. One can make a strong case that it's actually the cutting edge of the evolution of intelligence in general, because there's no indication that it's occurred anywhere else. To me that is what human civilization is all about. It is part of our destiny and part of the destiny of evolution to continue to progress ever faster, and to grow the power of intelligence exponentially. To contemplate stopping that—to think human beings are fine the way they are—is a misplaced fond remembrance of what human beings used to be. What human beings are is a species that has undergone a cultural and technological evolution, and it's the nature of evolution that it accelerates, and that its powers grow exponentially, and that's what we're talking about. The next stage of this will be to amplify our own intellectual powers with the results of our technology."

Published on KurzweilAI.net March 27, 2002. Originally published on http://www.edge.org March 25, 2002.

RAY KURZWEIL: My interest in the future really stems from my interest in being an inventor. I've had the idea of being an inventor since I was five years old, and I quickly realized that you had to have a good idea of the future if you're going to succeed as an inventor. It's a little bit like surfing; you have to catch a wave at the right time. I quickly realized the world quickly becomes a different place than it was when you started by the time you finally get something done. Most inventors fail not because they can't get something to work, but because all the market's enabling forces are not in place at the right time.

So I became a student of technology trends, and have developed mathematical models about how technology evolves in different areas like computers, electronics in general, communication storage devices, biological technologies like genetic scanning, reverse engineering of the human brain, miniaturization, the size of technology, and the pace of paradigm shifts. This helped guide me as an entrepreneur and as a technology creator so that I could catch the wave at the right time.

This interest in technology trends took on a life of its own, and I began to project some of them using what I call the law of accelerating returns, which I believe underlies technology evolution to future periods. I did that in a book I wrote in the 1980s, which had a road map of what the 1990s and the early 2000's would be like, and that worked out quite well. I've now refined these mathematical models, and have begun to really examine what the 21st century would be like. It allows me to be inventive with the technologies of the 21st century, because I have a conception of what technology, communications, the size of technology, and our knowledge of the human brain will be like in 2010, 2020, or 2030. If I can come up with scenarios using those

technologies, I can be inventive with the technologies of the future. I can't actually create these technologies yet, but I can write about them.

One thing I'd say is that if anything the future will be more remarkable than any of us can imagine, because although any of us can only apply so much imagination, there'll be thousands or millions of people using their imaginations to create new capabilities with these future technology powers. I've come to a view of the future that really doesn't stem from a preconceived notion, but really falls out of these models, which I believe are valid both for theoretical reasons and because they also match the empirical data of the 20th century.

One thing that observers don't fully recognize, and that a lot of otherwise thoughtful people fail to take into consideration adequately, is the fact that the pace of change itself has accelerated. Centuries ago people didn't think that the world was changing at all. Their grandparents had the same lives that they did, and they expected their grandchildren would do the same, and that expectation was largely fulfilled.

Today it's an axiom that life is changing and that technology is affecting the nature of society. But what's not fully understood is that the pace of change is itself accelerating, and the last 20 years are not a good guide to the next 20 years. We're doubling the paradigm shift rate, the rate of progress, every decade. So this will actually match the amount of progress we made in the whole 20th century, because we've been accelerating up to this point. The 20th century was like 25 years of change at today's rate of change. In the next 25 years we'll make four times the progress you saw in the 20th century. And we'll make 20,000 years of progress in the 21st century, which is almost a thousand times more technical change than we saw in the 20th century.

Specifically, computation is growing exponentially. The one exponential trend that people are aware of is called Moore's Law. But Moore's Law itself is just one method for bringing exponential growth to computers. People are aware that we're doubling the power of computation every 12 months because we can put twice as many transistors on an integrated circuit every two years. But in fact, they run twice as fast and double both the capacity and the speed, which means that the power quadruples.

What's not fully realized is that Moore's Law was not the first but the fifth paradigm to bring exponential growth to computers. We had electro-mechanical calculators, relay-based computers, vacuum tubes, and transistors. Every time one paradigm ran out of steam another took over. For a while there were shrinking vacuum tubes, and finally they couldn't make them any smaller and still keep the vacuum, so a whole different method came along. They weren't just tiny vacuum tubes, but transistors, which constitute a whole different approach. There's been a lot of discussion about Moore's Law running out of steam in about 12 years because by that time the transistors will only be a few atoms in width and we won't be able to shrink them any more. And that's true, so that particular paradigm will run out of steam.

We'll then go to the sixth paradigm, which is massively parallel computing in three dimensions. We live in a 3-dimensional world, and our brains organize in three dimensions, so we might as well compute in three dimensions. The brain processes information using an electrochemical

method that's ten million times slower than electronics. But it makes up for this by being three-dimensional. Every intra-neural connection computes simultaneously, so you have a hundred trillion things going on at the same time. And that's the direction we're going to go in. Right now, chips, even though they're very dense, are flat. Fifteen or twenty years from now computers will be massively parallel and will be based on biologically inspired models, which we will devise largely by understanding how the brain works.

We're already being significantly influenced by it. It's generally recognized, or at least accepted by a lot of observers, that we'll have the hardware to manipulate human intelligence within a brief period of time - I'd say about twenty years. A thousand dollars of computation will equal the 20 million billion calculations per second of the human brain. What's more controversial is whether or not we will have the software. People acknowledge that we'll have very fast computers that could in theory emulate the human brain, but we don't really know how the brain works, and we won't have the software, the methods, or the knowledge to create a human level of intelligence. Without this you just have an extremely fast calculator.

But our knowledge of how the brain works is also growing exponentially. The brain is not of infinite complexity. It's a very complex entity, and we're not going to achieve a total understanding through one simple breakthrough, but we're further along in understanding the principles of operation of the human brain than most people realize. The technology for scanning the human brain is growing exponentially, our ability to actually see the internal connection patterns is growing, and we're developing more and more detailed mathematical models of biological neurons. We actually have very detailed mathematical models of several dozen regions of the human brain and how they work, and have recreated their methodologies using conventional computation. The results of those re-engineered or re-implemented synthetic models of those brain regions match the human brain very closely.

We're also literally replacing sections of the brain that are degraded or don't work any more because of disabilities or disease. There are neural implants for Parkinson's Disease and well-known cochlear implants for deafness. There's a new generation of those that are coming out now that provide a thousand points of frequency resolution and will allow deaf people to hear music for the first time. The Parkinson's implant actually replaces the cortical neurons themselves that are destroyed by that disease. So we've shown that it's feasible to understand regions of the human brain, and reimplement those regions in conventional electronics computation that will actually interact with the brain and perform those functions.

If you follow this work and work out the mathematics of it. It's a conservative scenario to say that within 30 years - possibly much sooner - we will have a complete map of the human brain, we will have complete mathematical models of how each region works, and we will be able to re-implement the methods of the human brain, which are quite different than many of the methods used in contemporary artificial intelligence.

But these are actually similar to methods that I use in my own field—pattern recognition—which is the fundamental capability of the human brain. We can't think fast enough to logically analyze situations very quickly, so we rely on our powers of pattern recognition. Within 30 years we'll be able to create non-biological intelligence that's comparable to human intelligence. Just like a

biological system, we'll have to provide it an education, but here we can bring to bear some of the advantages of machine intelligence: Machines are much faster, and much more accurate. A thousand-dollar computer can remember billions of things accurately—we're hard-pressed to remember a handful of phone numbers.

Once they learn something, machines can also share their knowledge with other machines. We don't have quick downloading ports at the level of our intra-neuronal connection patterns and our concentrations of neurotransmitters, so we can't just download knowledge. I can't just take my knowledge of French and download it to you, but machines can. So we can educate machines through a process that can be hundreds or thousands of times faster than the comparable process in humans. It can provide a 20-year education to a human-level machine in maybe a few weeks or a few days and then these machines can share their knowledge.

The primary implication of all this will be to enhance our own human intelligence. We're going to be putting these machines inside our own brains. We're starting to do that now with people who have severe medical problems and disabilities, but ultimately we'll all be doing this. Without surgery, we'll be able to introduce calculating machines into the blood stream that will be able to pass through the capillaries of the brain. These intelligent, blood-cell-sized nanobots will actually be able to go to the brain and interact with biological neurons. The basic feasibility of this has already been demonstrated in animals.

One application of sending billions of nanobots into the brain is full-immersion virtual reality. If you want to be in real reality, the nanobots sit there and do nothing, but if you want to go into virtual reality, the nanobots shut down the signals coming from my real senses, replace them with the signals I would be receiving if I were in the virtual environment, and then my brain feels as if it's in the virtual environment. And you can go there yourself—or, more interestingly you can go there with other people—and you can have everything from sexual and sensual encounters to business negotiations, in full-immersion virtual reality environments that incorporate all of the senses.

People will beam their own flow of sensory experiences and the neurological correlates of their emotions out into the Web, the way people now beam images from web cams in their living rooms and bedrooms. This will enable you to plug in and actually experience what it's like to be someone else, including their emotional reactions, à la the plot concept of Being John Malkovich. In virtual reality you don't have to be the same person. You can be someone else, and can project yourself as a different person.

Most importantly, we'll be able to enhance our biological intelligence with non-biological intelligence through intimate connections. This won't mean just having one thin pipe between the brain and a non-biological system, but actually having non-biological intelligence in billions of different places in the brain. I don't know about you, but there are lots of books I'd like to read and Web sites I'd like to go to, and I find my bandwidth limiting. So instead of having a mere hundred trillion connections, we'll have a hundred trillion times a million. We'll be able to enhance our cognitive pattern recognition capabilities greatly, think faster, and download knowledge.

If you follow these trends further, you get to a point where change is happening so rapidly that there appears to be a rupture in the fabric of human history. Some people have referred to this as the "Singularity." There are many different definitions of the Singularity, a term borrowed from physics, which means an actual point of infinite density and energy that's kind of a rupture in the fabric of space-time.

Here, that concept is applied by analogy to human history, where we see a point where this rate of technological progress will be so rapid that it appears to be a rupture in the fabric of human history. It's impossible in physics to see beyond a Singularity, which creates an event boundary, and some people have hypothesized that it will be impossible to characterize human life after the Singularity. My question is, what will human life be like after the Singularity, which I predict will occur somewhere right before the middle of the 21st century?

A lot of the concepts we have of the nature of human life—such as longevity—suggest a limited capability as biological, thinking entities. All of these concepts are going to undergo significant change as we basically merge with our technology. It's taken me a while to get my own mental arms around these issues. In the book I wrote in the 1980s, The Age of Intelligent Machines, I ended with the specter of machines matching human intelligence somewhere between 2020 and 2050, and I basically have not changed my view on that time frame, although I left behind my view that this is a final specter. In the book I wrote ten years later, The Age of Spiritual Machines, I began to consider what life would be like past the point where machines could compete with us. Now I'm trying to consider what that will mean for human society.

One thing that we should keep in mind is that innate biological intelligence is fixed. We have 1026 calculations per second in the whole human race and there are ten billion human minds. Fifty years from now, the biological intelligence of humanity will still be at that same order of magnitude. On the other hand, machine intelligence is growing exponentially, and today it's a million times less than that biological figure. So although it still seems that human intelligence is dominating, which it is, the crossover point is around 2030 and non-biological intelligence will continue its exponential rise.

EDGE: This reminds me of a conversation I once had with John Lilly about dolphins. I asked him, "How do you know they're more intelligent than we are?" Isn't knowledge tautological? How can we know more than we do know? Who would know it, except us?

KURZWEIL: That's actually a very good point, because one response is not to want to be enhanced, not to have nanobots. A lot of people say that they just want to stay a biological person. But what will the Singularity look like to people who want to remain biological? The answer is that they really won't notice it, except for the fact that machine intelligence will appear to biological humanity to be their transcendent servants. It will appear that these machines are very friendly are taking care of all of our needs, and are really our transcendent servants. But providing that service of meeting all of the material and emotional needs of biological humanity will comprise a very tiny fraction of the mental output of the non-biological component of our civilization. So there's a lot that, in fact, biological humanity won't actually notice.

There are two levels of consideration here. On the economic level, mental output will be the primary criterion. We're already getting close to the point that the only thing that has value is information. Information has value to the extent that it really reflects knowledge, not just raw data. There are a few products on this table—a clock, a camera, tape recorder - that are physical objects, but really the value of them is in the information that went into their design: the design of their chips and the software that's used to invent and manufacture them. The actual raw materials - a bunch of sand and some metals and so on—is worth a few pennies, but these products have value because of all the knowledge that went into creating them.

And the knowledge component of products and services is asymptoting towards 100 percent. By the time we get to 2030 it will be basically 100 percent. With a combination of nanotechnology and artificial intelligence, we'll be able to create virtually any physical product and meet all of our material needs. When everything is software and information, it'll be a matter of just downloading the right software, and we're already getting pretty close to that.

On a spiritual level, the issue of what is consciousness is another important aspect of this, because we will have entities by 2030 that seem to be conscious, and that will claim to have feelings. We have entities today, like characters in your kids' video games, that can make that claim, but they are not very convincing. If you run into a character in a video game and it talks about its feelings, you know it's just a machine simulation; you're not convinced that it's a real person there. This is because that entity, which is a software entity, is still a million times simpler than the human brain.

In 2030, that won't be the case. Say you encounter another person in virtual reality that looks just like a human but there's actually no biological human behind it—it's completely an AI projecting a human-like figure in virtual reality, or even a human-like image in real reality using an android robotic technology. These entities will seem human. They won't be a million times simpler than humans. They'll be as complex as humans. They'll have all the subtle cues of being humans. They'll be able to sit here and be interviewed and be just as convincing as a human, just as complex, just as interesting. And when they claim to have been angry or happy it'll be just as convincing as when another human makes those claims.

At this point, it becomes a really deeply philosophical issue. Is that just a very clever simulation that's good enough to trick you, or is it really conscious in the way that we assume other people are? In my view there's no real way to test that scientifically. There's no machine you can slide the entity into where a green light goes on and says okay, this entity's conscious, but no, this one's not. You could make a machine, but it will have philosophical assumptions built into it. Some philosophers will say that unless it's squirting impulses through biological neurotransmitters, it's not conscious, or that unless it's a biological human with a biological mother and father it's not conscious. But it becomes a matter of philosophical debate. It's not scientifically resolvable.

The next big revolution that's going to affect us right away is biological technology, because we've merged biological knowledge with information processing. We are in the early stages of understanding life processes and disease processes by understanding the genome and how the genome expresses itself in protein. And we're going to find—and this has been apparent all

along—that there's a slippery slope and no clear definition of where life begins. Both sides of the abortion debate have been afraid to get off the edges of that debate: that life starts at conception on the one hand or it starts literally at birth on the other. They don't want to get off those edges, because they realize it's just a completely slippery slope from one end to the other.

But we're going to make it even more slippery. We'll be able to create stem cells without ever actually going through the fertilized egg. What's the difference between a skin cell, which has all the genes, and a fertilized egg? The only differences are some proteins in the eggs and some signaling factors that we don't fully understand, yet that are basically proteins. We will get to the point where we'll be able to take some protein mix, which is just a bunch of chemicals and clearly not a human being, and add it to a skin cell to create a fertilized egg that we can then immediately differentiate into any cell of the body. When I go like this and brush off thousands of skin cells, I will be destroying thousands of potential people. There's not going to be any clear boundary.

This is another way of saying also that science and technology are going to find a way around the controversy. In the future, we'll be able to do therapeutic cloning, which is a very important technology that completely avoids the concept of the fetus. We'll be able to take skin cells and create, pretty directly without ever going through a fetus, all the cells we need.

We're not that far away from being able to create new cells. For example, I'm 53 but with my DNA, I'll be able to create the heart cells of a 25-year-old man, and I can replace my heart with those cells without surgery just by sending them through my blood stream. They'll take up residence in the heart, so at first I'll have a heart that's one percent young cells and 99 percent older ones. But if I keep doing this every day, a year later, my heart is 99 percent young cells. With that kind of therapy we can ultimately replenish all the cell tissues and the organs in the body. This is not something that will happen tomorrow, but these are the kinds of revolutionary processes we're on the verge of.

If you look at human longevity—which is another one of these exponential trends—you'll notice that we added a few days every year to the human life expectancy in the 18th century. In the 19th century we added a few weeks every year, and now we're now adding over a hundred days a year, through all of these developments, which are going to continue to accelerate. Many knowledgeable observers, including myself, feel that within ten years we'll be adding more than a year every year to life expectancy.

As we get older, human life expectancy will actually move out at a faster rate than we're actually progressing in age, so if we can hang in there, our generation is right on the edge. We have to watch our health the old-fashioned way for a while longer so we're not the last generation to die prematurely. But if you look at our kids, by the time they're 20, 30, 40 years old, these technologies will be so advanced that human life expectancy will be pushed way out.

There is also the more fundamental issue of whether or not ethical debates are going to stop the developments that I'm talking about. It's all very good to have these mathematical models and these trends, but the question is if they going to hit a wall because people, for one reason or

another—through war or ethical debates such as the stem cell issue controversy—thwart this ongoing exponential development.

I strongly believe that's not the case. These ethical debates are like stones in a stream. The water runs around them. You haven't seen any of these biological technologies held up for one week by any of these debates. To some extent, they may have to find some other ways around some of the limitations, but there are so many developments going on. There are dozens of very exciting ideas about how to use genomic information and proteomic information. Although the controversies may attach themselves to one idea here or there, there's such a river of advances. The concept of technological advance is so deeply ingrained in our society that it's an enormous imperative. Bill Joy has gotten around—correctly—talking about the dangers, and I agree that the dangers are there, but you can't stop ongoing development.

The kinds of scenarios I'm talking about 20 or 30 years from now are not being developed because there's one laboratory that's sitting there creating a human-level intelligence in a machine. They're happening because it's the inevitable end result of thousands of little steps. Each little step is conservative, not radical, and makes perfect sense. Each one is just the next generation of some company's products. If you take thousands of those little steps—which are getting faster and faster—you end up with some remarkable changes 10, 20, or 30 years from now. You don't see Sun Microsystems saying the future implication of these technologies is so dangerous that they're going to stop creating more intelligent networks and more powerful computers. Sun can't do that. No company can do that because it would be out of business. There's enormous economic imperative.

There is also a tremendous moral imperative. We still have not millions but billions of people who are suffering from disease and poverty, and we have the opportunity to overcome those problems through these technological advances. You can't tell the millions of people who are suffering from cancer that we're really on the verge of great breakthroughs that will save millions of lives from cancer, but we're canceling all that because the terrorists might use that same knowledge to create a bioengineered pathogen.

This is a true and valid concern, but we're not going to do that. There's a tremendous belief in society in the benefits of continued economic and technological advance. Still, it does raise the question of the dangers of these technologies, and we can talk about that as well, because that's also a valid concern.

Another aspect of all of these changes is that they force us to re-evaluate our concept of what it means to be human. There is a common viewpoint that reacts against the advance of technology and its implications for humanity. The objection goes like this: we'll have very powerful computers but we haven't solved the software problem. And because the software's so incredibly complex, we can't manage it.

I address this objection by saying that the software required to emulate human intelligence is actually not beyond our current capability. We have to use different techniques - different self-organizing methods - that are biologically inspired. The brain is complicated but it's not that complicated. You have to keep in mind that it is characterized by a genome of only 23 million

bytes. The genome is six billion bits—that's eight hundred million bytes—and there are massive redundancies. One pretty long sequence called ALU is repeated 300 thousand times. If you use conventional data compression on the genomes (at 23 million bytes, a small fraction of the size of Microsoft Word), it's a level of complexity that we can handle. But we don't have that information yet.

You might wonder how something with 23 million bytes can create a human brain that's a million times more complicated than itself. That's not hard to understand. The genome creates a process of wiring a region of the human brain involving a lot of randomness. Then, when the fetus becomes a baby and interacts with a very complicated world, there's an evolutionary process within the brain in which a lot of the connections die out, others get reinforced, and it self-organizes to represent knowledge about the brain. It's a very clever system, and we don't understand it yet, but we will, because it's not a level of complexity beyond what we're capable of engineering.

In my view there is something special about human beings that's different from what we see in any of the other animals. By happenstance of evolution we were the first species to be able to create technology. Actually there were others, but we are the only one that survived in this ecological niche. But we combined a rational faculty, the ability to think logically, to create abstractions, to create models of the world in our own minds, and to manipulate the world. We have opposable thumbs so that we can create technology, but technology is not just tools. Other animals have used primitive tools, but the difference is actually a body of knowledge that changes and evolves itself from generation to generation. The knowledge that the human species has is another one of those exponential trends.

We use one stage of technology to create the next stage, which is why technology accelerates, why it grows in power. Today, for example, a computer designer has these tremendously powerful computer system design tools to create computers, so in a couple of days they can create a very complex system and it can all be worked out very quickly. The first computer designers had to actually draw them all out in pen on paper. Each generation of tools creates the power to create the next generation.

So technology itself is an exponential, evolutionary process that is a continuation of the biological evolution that created humanity in the first place. Biological evolution itself evolved in an exponential manner. Each stage created more powerful tools for the next, so when biological evolution created DNA it now had a means of keeping records of its experiments so evolution could proceed more quickly. Because of this, the Cambrian explosion only lasted a few tens of millions of years, whereas the first stage of creating DNA and primitive cells took billions of years. Finally, biological evolution created a species that could manipulate its environment and had some rational faculties, and now the cutting edge of evolution actually changed from biological evolution into something carried out by one of its own creations, Homo sapiens, and is represented by technology. In the next epoch this species that ushered in its own evolutionary process—that is, its own cultural and technological evolution, as no other species has—will combine with its own creation and will merge with its technology. At some level that's already happening, even if most of us don't necessarily have them yet inside our bodies and

brains, since we're very intimate with the technology—it's in our pockets. We've certainly expanded the power of the mind of the human civilization through the power of its technology.

We are entering a new era. I call it "the Singularity." It's a merger between human intelligence and machine intelligence that is going to create something bigger than itself. It's the cutting edge of evolution on our planet. One can make a strong case that it's actually the cutting edge of the evolution of intelligence in general, because there's no indication that it's occurred anywhere else. To me that is what human civilization is all about. It is part of our destiny and part of the destiny of evolution to continue to progress ever faster, and to grow the power of intelligence exponentially. To contemplate stopping that—to think human beings are fine the way they are—is a misplaced fond remembrance of what human beings used to be. What human beings are is a species that has undergone a cultural and technological evolution, and it's the nature of evolution that it accelerates, and that its powers grow exponentially, and that's what we're talking about. The next stage of this will be to amplify our own intellectual powers with the results of our technology.

What is unique about human beings is our ability to create abstract models and to use these mental models to understand the world and do something about it. These mental models have become more and more sophisticated, and by becoming embedded in technology, they have become very elaborate and very powerful. Now we can actually understand our own minds. This ability to scale up the power of our own civilization is what's unique about human beings.

Patterns are the fundamental ontological reality, because they are what persists, not anything physical. Take myself, Ray Kurzweil. What is Ray Kurzweil? Is it this stuff here? Well, this stuff changes very quickly. Some of our cells turn over in a matter of days. Even our skeleton, which you think probably lasts forever because we find skeletons that are centuries old, changes over within a year. Many of our neurons change over. But more importantly, the particles making up the cells change over even more quickly, so even if a particular cell is still there the particles are different. So I'm not the same stuff, the same collection of atoms and molecules that I was a year ago.

But what does persist is that pattern. The pattern evolves slowly, but the pattern persists. So we're kind of like the pattern that water makes in a stream; you put a rock in there and you'll see a little pattern. The water is changing every few milliseconds; if you come a second later, it's completely different water molecules, but the pattern persists. Patterns are what have resonance. Ideas are patterns, technology is patterns. Even our basic existence as people is nothing but a pattern. Pattern recognition is the heart of human intelligence. Ninety-nine percent of our intelligence is our ability to recognize patterns.

There's been a sea change just in the last several years in the public understanding of the acceleration of change and the potential impact of all of these technologies - computer technology, communications, biological technology - on human society. There's really been tremendous change in popular public perception in the past three years because of the onslaught

of stories and news developments that document and support this vision. There are now several stories every day that are significant developments and that show the escalating power of these technologies.

http://www.edge.org/3rd_culture/ kurzweil_singularity/kurzweil_singularity_index.html
(Audio and video available)

Accelerating Intelligence: Where Will Technology Lead Us?

Ray Kurzweil

http://www.kurzweilai.net/articles/art0450.html

Ray Kurzweil gave a Special Address at BusinessWeek's The Digital Economy New Priorities: Building A Collaborative Enterprise In Uncertain Times conference on December 6, 2001 in San Francisco. He introduced business CEOs to the Singularity — the moment when distinctions between human and machine intelligence disappear.

Published on KurzweilAI.net March 26, 2002

A video of the talk is now online at:
http://www.kurzweilai.net/articles/videos/ray_bizweek.ram

Excerpts from Ray Kurzweil's remarks:

"In considering the genesis of Moore's Law, I put 49 famous computing devices over the past century on an exponential graph. From this exercise, it became apparent that the acceleration of computing power did not start with integrated circuits, but has continued through multiple paradigm shifts (electromechanical calculators, relays, vacuum tubes, transistors, and finally integrated circuits).

"Moore's Law was not the first, but the fifth paradigm, to provide exponential growth in computing. The next paradigm, which will involve computing in three dimensions rather than the two manifested in today's flat chips, will lead to computing at the molecular, and ultimately the subatomic level. We can be confident that the acceleration of computing will survive the well-anticipated demise of Moore's Law.

"There are comparable exponential trends underlying a wide variety of other technologies: communications (both wired and wireless), brain scanning speeds and resolutions, genome scanning, and miniaturization (we are currently shrinking technology at a rate of 5.6 per linear dimension per decade). Even the rate of technological progress is speeding up, now doubling each decade. The mathematical models I've developed over the past couple of decades to describe these trends, which I call the law of accelerating returns, has proven predictive of the developments we've seen during the 1990s. From these models, I believe we can be confident of continued exponential growth in these and other technologies for the foreseeable future.

"By 2009, computers will disappear. Displays will be written directly onto our retinas by devices in our eyeglasses and contact lenses. In addition to virtual high-resolution displays, these intimate displays will provide full immersion visual virtual reality. We will have ubiquitous, very-high-bandwidth wireless connection to the Internet at all times. "Going to a web site" will mean entering a virtual reality environment — at least for the visual and auditory sense — where we will meet other real people. There will be simulated people as well, but the virtual

personalities will not be up to human standards, at least not by 2009. The electronics for all of this will be so small that it will be invisibly embedded in our glasses and clothing.

"By 2029, as a result of continuing trends in miniaturization, computation, and communication, we will have billions of nanobots — intelligent robots the same of blood cells or smaller — traveling through the capillaries of our brain communicating directly with our biological neurons. By taking up positions next to every nerve fiber coming from all of our senses, the nanobots will provide full-immersion virtual reality involving all five of the senses. So we will enter virtual reality environments (via the web, of course) of our choice and meet people, both real and virtual, only now the difference won't be so clear.

"Just as people today beam their images from little web cams out onto the Internet for others to share, many people in 2029 will beam the full stream of signals coming directly from their senses onto the web. We will then be able to experience what other people are experiencing, à la John Malkovich. Of course, the everyday lives of many such experience beamers may not be all that compelling, so there will be plenty of prerecorded experiences we can plug into it. Beyond just the five senses, these shared experiences will include emotional responses, sexual pleasure, and other mental reactions.

"Brain implants based on these distributed intelligent nanobots will extend our brains in every conceivable way, massively expanding our memory and otherwise vastly improving all of our sensory, pattern-recognition and cognitive abilities.

"Oh, and one more thing: we'll live a long time too. The expanding human life span is another one of those exponential trends. In the eighteenth century, we added a few days every year to human longevity; during the nineteenth century we added a couple of weeks each year; and now we're adding almost a half a year every year. With the revolutions in rational drug design, genomics, therapeutic cloning of our own organs and tissues, and related developments in bio-information sciences, we will be adding more than a year every year within ten years. So take care of yourself the old-fashioned way for just a little while longer, and you may actually get to experience the remarkable century ahead."

Max More and Ray Kurzweil on the Singularity

Ray Kurzweil, Max More

http://www.kurzweilai.net/articles/art0408.html

As technology accelerates over the next few decades and machines achieve superintelligence, we will encounter a dramatic phase transition: the "Singularity." Will it be a "wall" (a barrier as conceptually impenetrable as the event horizon of a black hole in space), an "AI-Singularity" ruled by super-intelligent AIs, or a gentler "surge" into a posthuman era of agelessness and super-intelligence? Will this meme be hijacked by religious "passive singularitarians" obsessed with a future rapture? Ray Kurzweil and Extropy Institute president Max More debate.

Published on KurzweilAI.net February 26, 2002.

Ray: You were one of the earliest pioneers in articulating and exploring issues of the acceleration of technology and transhumanism. What led you to this examination?

Max: This short question is actually an enormous question. A well-rounded answer would take longer than I dare impose on any reader! One short answer is this:

Before my interest in examining accelerating technological progress and issues of transhumanism, I first had a *realization*. I saw very clearly how limited are human beings in their wisdom, in their intellectual and emotional development, and in their sensory and physical capabilities. I have always felt dissatisfied with those limitations and faults. After an early-teens interest in what I'll loosely call (with mild embarrassment) "psychic stuff," as I came to learn more science and critical thinking, I ceased to give any credence to psychic phenomena, as well as to any traditional religious views. With those paths to any form of transcendence closed, I realized that transhumanity (as I began to think of it), would only be achieved through science and technology steered by human values.

So, the realization was in two parts: A recognition of the undesirable limitations of human nature. And an understanding that science and technology were essential keys to overcoming human nature's confines. In my readings in science, especially potent in encouraging me to think in terms of the development of intelligence rather than static Platonic forms was evolutionary theory. When I taught basic evolutionary theory to college students, I invariably found that about 95% of them had never studied it in school. Most quickly developed some understanding of it in the class, but some found it hard to adjust to a different perspective with so many implications. To me evolutionary thinking seemed natural. It only made it clearer that humanity need not be the pinnacle of evolution.

My drive to understand the issues in a transhumanist manner resulted from a melding of technological progress and philosophical perspective. Even before studying philosophy in the strict sense, I had the same essential worldview that included perpetual progress, practical optimism, and self-transformation. Watching the Apollo 11 moon landing at the age of 5, then all the Apollo launches to follow, excited me tremendously. At that time, space was the frontier to

explore. Soon, even before I had finished growing, I realized that the major barrier to crack first was that of human aging and mortality. In addition to tracking progress, from my early teens I started taking whatever reasonable measures I could to extend my life expectancy.

Philosophically, I formed an extropian/transhumanist perspective by incorporating numerous ideas and influences into what many have found to be a coherent framework of a perspective. Despite disagreeing with much (and not having read all) of Nietzsche's work, I do have a fondness for certain of his views and the way he expressed them. Most centrally, as a transhumanist, I resonate to Nietzsche's declaration that "Man is a rope, fastened between animal and overman—a rope over an abyss... What is great in man is that he is a bridge and not a goal."

A bridge, not a goal. That nicely summarizes a transhumanist perspective. We are not perfect. Neither are we to be despised or pitied or to debase ourselves before imaginary perfect beings. We are to see ourselves as a work in progress. Through ambition, intelligence, and a dash of good sense, we will progress from human to something better (according to my values).

Many others have influenced my interest in these ideas. Though not the deepest or clearest thinker, Timothy Leary's SMI^{2}LE (Space Migration, Intelligence Increase, Life Extension) formula still appeals to me. However, today I find issues such as achieving superlongevity, superintelligence, and self-sculpting abilities to be more urgent. After my earlier years of interest, I particularly grew my thinking by reading people including philosophers Paul Churchland and Daniel Dennett, biologist Richard Dawkins, Hans Moravec, Roy Walford, Marvin Minsky, Vernor Vinge, and most recently Ray Kurzweil, who I think has brought a delightful clarity to many transhumanist issues.

Ray: How do you define the Singularity?

Max: I believe the term "Singularity," as we are using it these days, was popularized by Vernor Vinge in his 1986 novel *Marooned in Realtime*. (It appears that the term was first used in something like this sense, but not implying superhuman intelligence, by John von Neumann in the 1950s.) Vinge's own usage seems to leave an exact definition open to varying interpretations. Certainly it involves an accelerating increase in machine intelligence culminating in a sudden shift to super intelligence, either through the awakening of networked intelligence or the development of individual AIs. From the human point of view, according to Vinge, this change "will be a throwing away of all the previous rules, perhaps in the blink of an eye." Since the term means different things to different people, I will give three definitions.

Singularity #1: This Singularity includes the notion of a "wall" or "prediction horizon"—a time horizon beyond which we can no longer say anything useful about the future. The pace of change is so rapid and deep that our human minds cannot sensibly conceive of life post-Singularity. Many regard this as a specific point in time in the future, sometimes estimated at around 2035 when AI and nanotechnology are projected to be in full force. However, the prediction-horizon definition does not require such an assumption. The more that progress accelerates, the shorter the distance measured in years that we may see ahead. But as we progress, the prediction horizon, while probably shortening in time, will also move further out. So this definition could be broken into two, one of which insists on a particular date for a prediction horizon, while the

other acknowledges a moving horizon. One argument for assigning a point in time is based on the view that the emergence of super-intelligence will be a singular advance, an instantaneous break with all the rules of the past.

Singularity #2: We might call this the AI-Singularity, or Moravec's Singularity since it most closely resembles the detailed vision of roboticist Hans Moravec. In this Singularity humans have no guaranteed place. The Singularity is driven by super-intelligent AI, which immediately follows from human-level AI. Without the legacy hardware of humans, these AIs leave humans behind in a runaway acceleration. In some happier versions of this type of Singularity, the super-intelligent AIs benevolently "uplift" humans to their level by means of brain uploading.

Singularity #3: Singularity seen as a *surge* into a transhuman and posthuman era. This view, though different in its emphasis, is compatible with the shifting time-horizon version of Singularity #1. In Singularity as Surge the rate of change need not remotely approach infinity (as a mathematical singularity). In this view, technological progress will continue to accelerate, though perhaps not quite as fast as some projections suggest, rapidly but not discontinuously transforming the human condition.

This could be termed a Singularity for two reasons: First, it would be a historically brief phase transition from the human condition to a posthuman condition of agelessness, super-intelligence, and physical, intellectual, and emotional self-sculpting. This dramatic phase transition, while not mathematically instantaneous, will mean an unprecedented break from the past. Second, since the posthuman condition (itself continually evolving) will be so radically different from human life, it will likely be largely if not completely incomprehensible to humans as we are today. Unlike some versions of the Singularity, the Surge/phase transition view allows that people may be at different stages along the path to posthuman at the same time, and that we may become posthuman in stages rather than all at once. For instance, I think it fairly likely that we achieve superlongevity before super-intelligence.

Ray: Do you see a Singularity in the future of human civilization?

Max: I do see a Singularity of the third kind in our future. A historically, if not subjectively, extremely fast phase change from human to transhuman to posthuman appears as a highly likely scenario. I do not see it as inevitable. It will take vast amounts of hard work, intelligence, determination, and some wisdom and luck to achieve. It's possible that some humans will destroy the race through means such as biological warfare. Or our culture may rebel against change, seduced by religious and cultural urgings for "stability," "peace" and against "hubris" and "the unknown,"

Although a Singularity as Surge could be stopped or slowed in these and other ways (massive meteorite strike?), I see the most likely scenario as being a posthuman Singularity. This is strongly implied by current accelerating progress in numerous fields including computation, materials science, bioinformatics and the convergence of infotech neuroscience, and biotech, microtech and nanotech.

Although I do not see super-intelligence alone as the *only* aspect of a Singularity, I do see it as a central aspect and driver. I grant that it is entirely possible that super-intelligence will arrive in the form of a *deus ex machina*, a runaway single-AI super-intelligence. However, my tentative assessment suggests that the Singularity is more likely to arise from one of two other means suggested by Vinge in his 1993 essay[1]. It could result from large computer networks of computers and their users—some future version of a semantic Web—"waking up" in the form of a distributed super-intelligent entity or community of minds. It could also (not exclusive of the previous scenario) result from increasingly intimate human-computer interfaces (by "computer," I loosely include all manner of sensors, processors, and networks). At least in the early stages, and partly in combination with human-computer interfaces, I expect biological human intelligence to be augmented through the biological sciences.

To summarize: I do not expect an *instantaneous* Singularity, nor one in which humans play no part after the creation of a self-improving human-level AI. I do anticipate a Singularity in the form of a growing surge in the pace of change, leading to a transhuman transition. This phase change will be a historically rapid and deep change in the evolutionary process. This short period will put an end to evolution in thrall to our genes. Biology will become an increasingly vestigial component of our nature. Biological evolution will become ever more suffused with and replaced by technological evolution, until we pass into the posthuman era.

As a postscript to this answer, I want to sound a note of caution. As the near-universal prevalence of religious belief testifies, humans tend to attach themselves, without rational thought, to belief systems that promise some form of salvation, heaven, paradise, or nirvana. In the Western world, especially in millenarian Christianity, millions are attracted to the notion of sudden salvation and of a "rapture" in which the saved are taken away to a better place.

While I do anticipate a Singularity as Surge, I am concerned that the Singularity concept is especially prone to being hijacked by this memeset. This danger especially arises if the Singularity is thought of as occurring at a specific point in time, and even more if it is seen as an *inevitable* result of the work of others. I fear that many otherwise rational people will be tempted to see the Singularity as a form of salvation, making personal responsibility for the future unnecessary. Already, I see a distressing number of superlongevity advocates who apparently do not exercise or eat healthily, instead firmly *hoping* that medical technology will cure aging before they die. Clearly this abdication of personal responsibility is not inherent in the Singularity concept.

But I do see the concept as an attractor that will draw in those who treat it in this way. The only way I could see this as a good thing is if the Passive Singularitarians (as I will call them) substitute the Singularity for preexisting and much more unreasonable beliefs. I think those of us who speak of the Singularity should be wary of this risk if we value critical thought and personal responsibility. As much as I like Vernor and his thinking, I get concerned reading descriptions of the Singularity such as "a throwing away of all the previous rules, perhaps in the blinking of an eye." This comes dangerously close to encouraging a belief in a Future Rapture.

Ray: When will the Singularity take place?

Max: I cannot answer this question with any precision. I feel more confident predicting general trends than specific dates. Some trends look very clear and stable, such as the growth in computer power and storage density. But I see enough uncertainties (especially in the detailed understanding of human intelligence) in the breakthroughs needed to pass through a posthuman Singularity to make it impossible to give one date. Many Singularity exponents see several trends in computer power and atomic control over matter reaching critical thresholds around 2030. It does look like we will have computers with hardware as powerful as the human brain by then, but I remain to be convinced that this will immediately lead to superhuman intelligence. I also see a tendency in many projections to take a purely technical approach and to ignore possible economic, political, cultural, and psychological factors that could dampen the advances and their impact.

I will make one brief point to illustrate what I mean: Electronic computers have been around for over half a decade, and used in business for decades. Yet their effect on productivity and economic growth only became evident in the mid-1990s as corporate organization and business processes finally reformed to make better use of the new technology. Practice tends to lag technology, yet projections rarely allow for this. (This factor also led to many dot-com busts where business models required consumers to change their behavior in major ways.)

Cautions aside, I would be surprised (and, of course, disappointed) if we did not move well into a posthuman Singularity by the end of this century. I think that we are already at the very near edge of the transhuman transition. This will gather speed, and could lead to a Singularity as phase transition by the middle of the century. So, I will only be pinned down to the extent of placing the posthuman Singularity as not earlier than 2020 and probably not later than 2100, with a best *guess* somewhere around the middle of the century. Since that puts me in my 80s or 90s, I hope I am unduly pessimistic!

Ray: Thanks, Max, for these thoughtful and insightful replies. I appreciate your description of transhumanity as a transcendence to be "achieved through science and technology steered by human values." In this context, Nietzsche's "Man is a rope, fastened between animal and overman—a rope over an abyss" is quite pertinent, thereby interpreting Nietzsche's "overman" to be a reference to transhumanism.

The potential to hijack the concept of the Singularity by the "future rapture" memeset is discerning, but I would point out that humankind's innate inclination for salvation is not necessarily irrational. Perhaps we have this inclination precisely to anticipate the Singularity. Maybe it is the Singularity that has been hijacked by irrational belief systems, rather than the other way around. However, I share your antipathy toward passive singularitarianism. If technology is a double-edged sword, then there is the possibility of technology going awry as it surges toward the singularity to profoundly disturbing consequences. We do need to keep our eye on the ethical ball.

I don't agree that a cultural rebellion "seduced by religious and cultural urgings for 'stability,' peace,' and against 'hubris' and 'the unknown'" are likely to derail technological acceleration. Epochal events such as two world wars, the Cold War, and numerous economic, cultural, and social upheavals have failed to provide the slightest dent in the pace of the fundamental trends.

As I discuss below, recessions, including the Great Depression, register as only slight deviations from the far more profound effect of the underlying exponential growth of the economy, fueled by the exponential growth of information-based technologies.

The primary reason that technology accelerates is that each new stage provides more powerful tools to create the next stage. The same was true of the biological evolution that created the technology creating species in the first place. Indeed, each stage of evolution adds another level of indirection to its methods, and we can see technology itself as a level of indirection for the biological evolution that resulted in technological evolution.

To put this another way, the ethos of scientific and technological progress is so deeply ingrained in our civilization that halting this process is essentially impossible. Occasional ethical and legal impediments to fairly narrow developments are rather like stones in the river of advancement; progress just flows around them.

You wrote that it appears that we will have sufficient computer power to emulate the human brain by 2030, but that this development will not necessarily "immediately lead to superhuman intelligence." The other very important development is the accelerating process of reverse engineering (i.e., understanding the principles of operation of) the human brain. Without repeating my entire thesis here, this area of knowledge is also growing exponentially, and includes increasingly detailed mathematical models of specific neuron types, exponentially growing price-performance of human brain scanning, and increasingly accurate models of entire brain regions. Already, at least two dozen of the several hundred regions in the brain have been satisfactorily reverse-engineered and implemented in synthetic substrates. I believe that it is a conservative scenario to expect that the human brain will be fully reverse-engineered in a sufficiently detailed way to recreate its mental powers by 2029.

As you and I have discussed on various occasions, I've done a lot of thinking over the past few decades about the laws of technological evolution. As I mentioned above, technological evolution is a continuation of biological evolution. So the laws of technological evolution are compatible with the laws of evolution in general.

These laws imply at least a "surge" form of Singularity, as you describe it, during the first half of this century. This can be seen from the predictions for a wide variety of technological phenomena and indicators, including the power of computation, communication bandwidths, technology miniaturization, brain reverse engineering, the size of the economy, the rate of paradigm shift itself, and many others. These models apply to measures of both hardware and software.

Of course, if a mathematical line of inquiry yields counterintuitive results, then it makes sense to check the sensibility of these conclusions in another manner. However, in thinking through how the transformations of the Singularity will actually take place, through distinct periods of transhuman and posthuman development, the predictions of these formulae do make sense to me, in that one can describe each stage and how it is the likely effect of the stage preceding it.

To me, the concept of the Singularity as a "wall" implies a period of infinite change, that is, a mathematical Singularity. If there is a point in time at which change is infinite, then there is an inherent barrier in looking beyond this point in time. It becomes as impenetrable as the event horizon of a black hole in space, in which the density of matter and energy is infinite. The concept of the Singularity as a "surge," on the other hand, is compatible with the idea of exponential growth. It is the nature of an exponential function that it starts out slowly, then grows quite explosively as one passes what I call the "knee of the curve." From the surge perspective, the growth rate never becomes literally infinite, but it may appear that way from the limited perspective of someone who cannot follow such enormously rapid change.

This perspective can be consistent with the idea that you mention of the prediction horizon "moving further out," because one of the implications of the Singularity as a surge phenomenon is that humans will enhance themselves through intimate connection with technology, thereby increasing our capacity to understand change. So changes that we could not follow today may very well be comprehensible when they occur. I think that it is fair to say that at any point in time, the changes that occur will be comprehensible to *someone*, albeit the someone may be a superintelligence on the cutting edge of the Singularity.

With the above as an introduction, I thought you would find of interest a dialog that Hans Moravec and I had on the issue you address on whether the Singularity is a "wall" (i.e., "prediction horizon") or a "surge." Some of these ideas are best modeled in the language of mathematics, but I will try to put the math in a box, so to speak, as it is not necessary to track through the formulas in order to understand the basic ideas.

First, let me describe the formulas that I showed to Hans.

The following analysis describes the basis of understanding evolutionary change as an exponentially growing phenomenon (a double exponential to be exact). I will describe here the growth of computational power, although the formulas are similar for other exponentially growing aspects of information-based aspects of evolution, including our knowledge of human intelligence, which is a primary source of the software of intelligence.

We are concerned with three variables:

V: Velocity (i.e., power) of computation (measured in Calculations per Second per unit cost)

W: World Knowledge as it pertains to designing and building computational devices

t: Time

As a first-order analysis, we observe that computer power is a linear function of the knowledge of how to build computational devices. We also note that knowledge is cumulative, and that the instantaneous increment to knowledge is proportional to computational power. These observations result in the conclusion that computational power grows exponentially over time.

See Analysis One

The data that I've gathered shows that there is exponential growth in the rate of exponential growth (we doubled computer power every three years early in the 20th century, every two years in the middle of the century, and are doubling it every one year now).

The exponentially growing power of technology results in exponential growth of the economy. This can be observed going back at least a century. Interestingly, recessions, including the Great Depression, can be modeled as a fairly weak cycle on top of the underlying exponential growth. In each case, the economy "snaps back" to where it would have been had the recession/depression never existed in the first place. We can see even more rapid exponential growth in specific industries tied to the exponentially growing technologies, such as the computer industry.

If we factor in the exponentially growing resources for computation, we can see the source for the second level of exponential growth.

See Analysis Two

Now, let's consider some real-world data. My estimate of brain capacity is 100 billion neurons times an average 1,000 connections per neuron (with the calculations taking place primarily in the connections) times 200 calculations per second—a total of 20 million billion ($2*10^{16}$) calculations per second. Although these estimates are conservatively high, one can find higher and lower estimates. However, even much higher (or lower) estimates by orders of magnitude only shift the prediction by a relatively small number of years.

See Analysis Three

Human Brain = 100 Billion (10^{11}) neurons * 1000 (10^3) Connections/Neuron * 200 ($2 * 10^2$) Calculations Per Second Per Connection = $2 * 10^{16}$ Calculations Per Second

Human Race = 10 Billion (10^{10}) Human Brains = $2 * 10^{26}$ Calculations Per Second

We achieve one Human Brain capability ($2 * 10^{16}$ cps) for $1,000 around the year 2023.

We achieve one Human Brain capability ($2 * 10^{16}$ cps) for one cent around the year 2037.

We achieve one Human Race capability ($2 * 10^{26}$ cps) for $1,000 around the year 2049.

We achieve one Human Race capability ($2 * 10^{26}$ cps) for one cent around the year 2059.

If we factor in the exponentially growing economy, particularly with regard to the resources available for computation (already about a trillion dollars per year), we can see that nonbiological intelligence will be many trillions of times more powerful than biological intelligence by approximately mid-century.

Although the above analysis pertains to computational power, a comparable analysis can be made of brain reverse-engineering, i.e., knowledge about the principles of operation of human intelligence. There are many different ways to measure this, including mathematical models of

human neurons, the resolution, speed, and bandwidth of human brain scanning, and knowledge about the digital-controlled analog, massively parallel algorithms utilized in the human brain.

As I mentioned above, we have already succeeded in developing highly detailed models of several dozen of the several hundred regions of the brain and implementing these models in software, with very successful results. I won't describe all of this in our dialog here, but I will be reporting on brain reverse-engineering in some detail in my next book, *The Singularity is Near*.

We can view this effort as analogous to the genome project. The effort to understand the information processes in our biological heritage has largely completed the stage of collecting the raw genomic data, is now rapidly gathering the proteomic data, and has made a good start at understanding the methods underlying this information. With regard to the even more ambitious project to understand our neural organization, we are now approximately where the genome project was about ten years ago, but are further along than most people realize.

Keep in mind that the brain is the result of chaotic processes (which themselves use a controlled form of evolutionary pruning) described by a genome with very little data (only about 23 million bytes compressed). The analyses I will present in *The Singularity is Near* demonstrate that it is quite conservative to expect that we will have a complete understanding of the human brain and its methods, and thereby the software of human intelligence, prior to 2030.

The above is my own analysis, at least in mathematical terms, and backed up by extensive real-world data, of the Singularity as a "surge" phenomenon. This, then, is the conservative view of the Singularity.

Hans Moravec points out that my assumption that computer power grows proportionally with knowledge (i.e., V = C1 * W) is overly pessimistic because independent innovations (each of which is a linear increment to knowledge) increase the power of the technology in a multiplicative way, rather than an additive way.

In an email to me on February 15, 1999, Hans wrote:

"For instance, one (independent innovation) might be an algorithmic discovery (like log N sorting) that lets you get the same result with half the computation. Another might be a computer organization (like RISC) that lets you get twice the computation with the same number of gates. Another might be a circuit advance (like CMOS) that lets you get twice the gates in a given space. Others might be independent speed-increasing advances, like size-reducing copper interconnects and capacitance-reducing silicon-on-insulator channels. Each of those increments of knowledge more or less multiplies the effect of all of the others, and computation would grow exponentially in their number."

So if we substitute, as Hans suggests, V = exp(W) rather than V = C1 * W, then the result is that both W and V become infinite.

See Analysis Four

Hans and I then engaged in a dialog as to whether or not it is more accurate to say that computer power grows exponentially with knowledge (which is suggested by an analysis of independent innovations) (i.e., V = exp (W)), or grows linearly with knowledge (i.e., V = C1 * W) as I had originally suggested. We ended up agreeing that Hans' original statement that V = exp (W) is too "optimistic," and that my original statement that V = C1 * W is too "pessimistic."

We then looked at what is the "weakest" assumption that one could make that nonetheless results in a mathematical singularity. Hans wrote:

"But, of course, a lot of new knowledge steps on the toes of other knowledge, by making it obsolete, or diluting its effect, so the simple independent model doesn't work in general. Also, simply searching through an increasing amount of knowledge may take increasing amounts of computation. I played with the V=exp(W) assumption to weaken it, and observed that the singularity remains if you assume processing increases more slowly, for instance V = exp(sqrt(W)) or exp($W^{1/4}$). Only when V = exp(log(W)) (i.e., = W) does the progress curve subside to an exponential.

Actually, the singularity appears somewhere in the I-would-have-expected tame region between and V = W and V = W^2 (!)

Unfortunately the transitional territory between the merely exponential V=W and the singularity-causing V=W^2 is analytically hard to deal with. I assume just before a singularity appears, you get non-computably rapid growth!"

Interestingly, if we take my original assumption that computer power grows linearly with knowledge, but add that the resources for computation also grows in the same way, then the total amount of computational power grows as the square of knowledge, and again, we have a mathematical singularity.

See Analysis Five

The conclusions that I draw from these analyses are as follows. Even with the "conservative" assumptions, we find that nonbiological intelligence crosses the threshold of matching and then very quickly exceeds biological intelligence (both hardware and software) prior to 2030. We then note that nonbiological intelligence will then be able to combine the powers of biological intelligence with the ways in which nonbiological intelligence already excels, in terms of accuracy, speed, and the ability to instantly share knowledge.

Subsequent to the achievement of strong AI, human civilization will go through a period of increasing intimacy between biological and nonbiological intelligence, but this transhumanist period will be relatively brief before it yields to a posthumanist period in which nonbiological intelligence vastly exceeds the powers of unenhanced human intelligence. This, at least, is the conservative view.

The more "optimistic" view is difficult for me to imagine, so I assume that the formulas stop just short of the point at which the result becomes noncomputably large (i.e., infinite). What is interesting in the dialog that I had with Hans above is how easily the formulas can produce a

mathematical singularity. Thus the difference between the Singularity as "wall" and the Singularity as "surge" results from a rather subtle difference in our assumptions.

Max: Ray, I know I'm in good company when the "conservative" view means that we achieve superhuman intelligence and a posthuman transition by 2030! I suppose that puts me in the unaccustomed position of being the ultra-conservative. From the regular person's point of view, the differences between our expectations will seem trivial. Yet I think these critical comparisons are valuable in deciding whether the remaining human future is 25 years or 50 years or more. Differences in these estimations can have profound effects on outlook and which plans for the future are rational. One example would be the sense of saving heavily to build compound returns versus spending almost everything until the double exponential really turns strongly toward a vertical ascent.

I find your trend analysis compelling and certainly the most comprehensive and persuasive ever developed. Yet I am not quite willing to yield fully to the mathematical inevitability of your argument. History since the Enlightenment makes me wary of all arguments to inevitability, at least when they point to a specific time. Clearly your arguments are vastly more detailed and well-grounded than those of the 18th century proponents of inevitable progress. But I suspect that a range of non-computational factors could dampen the growth curve. The double exponential curve may describe very well the development of new technologies (at least those driven primarily by computation), but not necessarily their full implementation and effects.

Numerous world-changing technologies from steel mills to electricity to the telephone to the Internet have taken decades to move from introduction to widespread effects. We could point to the Web to argue that this lag between invention and full adoption is shrinking. I would agree for the most part, yet different examples may tell a different story. Fuel cells were invented decades ago but only now do they seem poised to make a major contribution to our energy supply.

Psychological and cultural factors act as future-shock absorbers. I am not sure that models based on models of evolution in information technology necessarily take these factors fully into account. In working with businesses to help them deal with change, I see over and over the struggle involved in altering organizational culture and business processes to take advantage of powerful software solutions from supply chain management to customer relationship management (CRM). CRM projects have a notoriously high failure rate, not because the software is faulty but because of poor planning and a failure to re-engineer business processes and employee incentives to fit.

I expect we will eventually reach a point where cognitive processes and emotions can be fully understood and modulated and where we have a deep understanding of social processes. These will then cease to act as significant brakes to progress. But major advances in those areas seem likely to come close to the Singularity and so will act as drags until very close. It could be that your math models may overstate early progress toward the Singularity due to these factors. They may also *understate* the last stages of progress as practice catches up to technology with the liberation of the brain from its historical limitations.

Apart from these human factors, I am concerned that other trends may not present such an optimistic picture of accelerating progress. Computer programming languages and tools have improved over the last few decades, but it seems they improve slowly. Yes, at some point computers will take over most programming and perhaps greatly accelerate the development of programming tools. Or humans will receive hippocampus augmentations to expand working memory. My point is not that we will not reach the Singularity but that different aspects of technology and humanity will advance at different rates, with the slower holding back the faster.

I have no way of formally modeling these potential braking factors, which is why I refrain from offering specific forecasts for a Singularity. Perhaps they will delay the transhuman transition by only a couple of years, or perhaps by 20. I would agree that as information technology suffuses ever more of economy and society, its powerful engines of change will accelerate everything faster and faster as time goes by. Therefore, although I am not sure that your equations will always hold, I do expect actual events to converge on your models the closer we get to Singularity.

I would like briefly to make two other comments on your reply. First, you suggest that "humankind's innate inclination for salvation is not necessarily irrational. Perhaps we have this inclination precisely to anticipate the Singularity." I am not sure how to take this suggestion. A natural reading suggests a teleological interpretation: humans have been given (genetically or culturally) this inclination. If so, who gave us this? Since I do not believe the evidence supports the idea that we are designed beings, I don't think such a teleological view of our inclination for salvation is plausible.

I would also say that I don't regard this inclination as inherently irrational. The inclination may be a side effect of the apparently universal human desire to understand and to solve problems. Those who feel helpless to solve certain kinds of problems often want to believe there is a higher power (tax accountant, car mechanic, government, or god) that can solve the problem. I would say such an inclination only becomes irrational when it takes the form of unfounded stories that are taken as literal, explanatory facts rather than symbolic expressions of deep yearnings. That aside, I am curious how you think that we come to have this inclination in order to anticipate the Singularity.

Second, you say that you don't agree that a cultural rebellion "seduced by religious and cultural urgings for 'stability,' peace,' and against 'hubris' and 'the unknown'" are "likely to derail technological acceleration." We really don't disagree here. If you look again at what I wrote, you can see that I do *not* think this derailing is *likely*. More exactly, while I think they are highly likely locally (look at the Middle East for example), they would have a hard time in today's world universally stopping or appreciably slowing technological progress. My concern was to challenge the idea that progress is inevitable rather than simply highly likely.

This point may seem unimportant if we adopt a position based on overall trends. But it will certainly matter to those left behind, temporarily or permanently in various parts of the world: The Muslim woman dying in childbirth as her culture refuses her medical attention; a dissident executed for speaking out against the state; or a patient who dies of nerve degeneration or who loses their personality due to amentia because religious conservatives have halted progress in

stem cell research. The derailing of progress is likely to be temporary and local, but no less real and potentially deadly for many. A more widespread and enduring throwback, perhaps due to a massively infectious and deadly terrorist attack, surely cannot be ruled out. Recent events have reminded us that the future needs security as well as research.

Normally I do the job of arguing that technological change will be faster than expected. Taking the other side in this dialog has been a stimulating change of pace. While I expect those major events we call the Singularity to come just a little later than you calculate, I strongly hope that I am mistaken and that you are correct. The sooner we master these technologies, the sooner we will conquer aging and death and all the evils that humankind has been heir to.

Ray: It's tempting indeed to continue this dialog indefinitely, or at least until the Singularity comes around. I am sure that we will do exactly that in a variety of forums. A few comments for the moment, however:

You cite the difference in our future perspective (regarding the time left until the Singularity) as being about 20 years. Of course, from the perspective of human history, let alone evolutionary history, that's not a very big difference. It's not clear that we differ by even that much. I've projected the date 2029 for a nonbiological intelligence to pass the Turing test. (As an aside, I just engaged in a "long term wager" with Mitchell Kapor to be administered by the "Long Now Foundation" on just this point.)

However, the threshold of a machine passing a valid Turing test, although unquestionably a singular milestone, does not represent the Singularity. This event will not immediately alter human identity in such a profound way as to represent the tear in the fabric of history that the term Singularity implies. It will take a while longer for all of these intertwined trends—biotechnology, nanotechnology, computing, communications, miniaturization, brain reverse engineering, virtual reality, and others—to fully mature. I estimate the Singularity at around 2045. You estimated the "posthuman Singularity as [occurring]. . . . with a best guess somewhere around the middle of the century." So, perhaps, our expectations are close to being about five years apart. Five years will in fact be rather significant in 2045, but even with our mutual high levels of impatience, I believe we will be able to wait that much longer

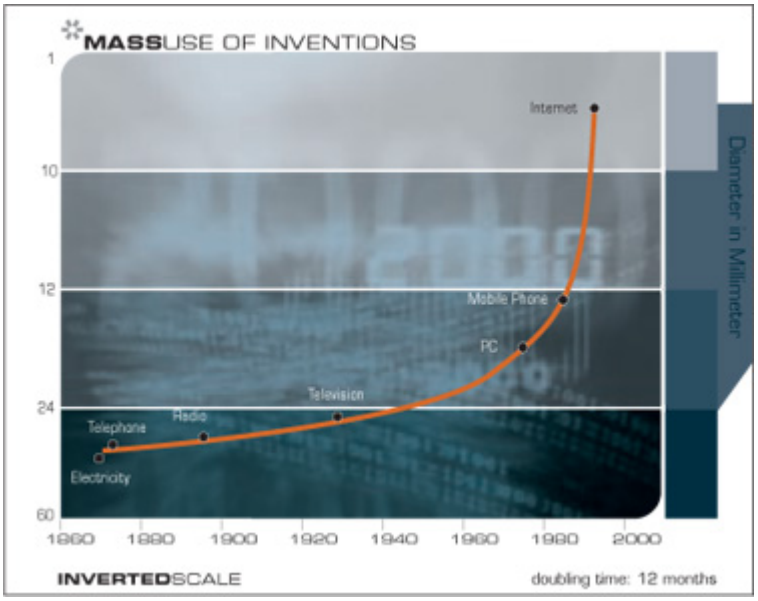
I do want to comment on your term "the remaining human future" (being "25 years or 50 years or more"). I would rather consider the post-Singularity period to be one that is "post-biological" rather than "posthuman." In my view, the other side of the Singularity may properly be considered still human and still infused with (our better) human values. At least that is what we need to strive for. The intelligence we are creating will be derived from human intelligence, i.e., derived from human designs, and from the reverse engineering of human intelligence.

As the beautiful images and descriptions that (your wife) Natasha Vita-More and her collaborators put together ("Radical Body Design 'Primo 3M+'") demonstrate, we will gladly move beyond the limitations, not to mention the pain and discomfort, of our biological bodies and brains. As William Butler Yeats wrote, an aging man's biological body is "but a paltry thing, a tattered coat upon a stick." Interestingly, Yeats concludes, "Once out of nature I shall never take, My bodily form from any natural thing, But such a form as Grecian goldsmiths make, Of

hammered gold and gold enamelling." I suppose that Yeats never read Feynman's treatise on nanotechnology or he would have mentioned carbon nanotubes.

I am concerned that if we refer to a "remaining human future," this terminology may encourage the perspective that something profound is being lost. I believe that a lot of the opposition to these emerging technologies stems from this uninformed view. I think we are in agreement that nothing of true value needs to be lost.

You write that "Numerous world-changing technologies. . . . have taken decades to move from introduction to widespread effects." But keep in mind that, in accordance with the law of exponential returns, technology adoption rates are accelerating along with everything else. The following chart shows the adoption time of various technologies, measured from invention to adoption by a quarter of the U.S. population:



With regard to technologies that are not information-based, the exponent of exponential growth is definitely slower than for computation and communications, but nonetheless positive, and as you point out, fuel cell technologies are posed for rapid growth. As one example of many, I'm involved with one company that has applied MEMS technology to fuel cells. Ultimately we will also see revolutionary changes in transportation from nanotechnology combined with new energy technologies (think microwings).

A common challenge to the feasibility of strong AI, and therefore to the Singularity, is to distinguish between quantitative and qualitative trends. This challenge says, in essence, that perhaps certain brute force capabilities such as memory capacity, processor speed, and communications bandwidths are expanding exponentially, but the qualitative aspects are not.

This is the hardware versus software challenge, and it is an important one. With regard to the price-performance of software, the comparisons in virtually every area are dramatic. Consider speech recognition software as one example of many. In 1985, $5,000 bought you a speech recognition software package that provided a 1,000 word vocabulary, did not provide continuous speech capability, required three hours of training, and had relatively poor accuracy. Today, for only $50, you can purchase a speech-recognition software package with a 100,000 word vocabulary, that does provide continuous speech capability, requires only five minutes of training, has dramatically improved accuracy, provides natural language understanding ability (for editing commands and other purposes), and many other features.

What about software development itself? I've been developing software myself for forty years, so I have some perspective on this. It's clear that the growth in productivity of software development has a lower exponent, but it is nonetheless growing exponentially. The development tools, class libraries, and support systems available today are dramatically more effective than those of decades ago. I have today small teams of just three or four people who achieve objectives in a few months that are comparable to what a team of a dozen or more people could accomplish in a year or more 25 years ago. I estimate the doubling time of software productivity to be approximately six years, which is slower than the doubling time for processor price-performance, which today is approximately one year. However, software productivity is nonetheless growing exponentially.

The most important point to be made here is that we have a specific game plan (i.e., brain reverse engineering) for achieving the software of human-level intelligence in a machine. It's actually not my view that brain reverse engineering is the only way to achieve strong AI, but this scenario does provide an effective existence-proof of a viable path to get there.

If you speak to some of the (thousands of) neurobiologists who are diligently creating detailed mathematical models of the hundreds of types of neurons found in the brain, or who are modeling the patterns of connections found in different regions, you will often encounter the common engineer's/scientist's myopia that results from being immersed in the specifics of one aspect of a large challenge. I'll discuss this challenge in some detail in the book I'm now working on, but I believe it is a conservative projection to expect that we will have detailed models of the several hundred regions of the brain within about 25 years (we already have impressively detailed models and simulations for a couple dozen such regions). As I alluded to above, only about half of the genome's 23 million bytes of useful information (i.e., what's left of the 800 million byte genome after compression) specifies the brain's initial conditions.

The other "non-computational factors [that] could dampen the growth curve" that you cite are "psychological and cultural factors [acting] as future-shock absorbers." You describe organizations you've worked with in which the "culture and business processes" resist change for a variety of reasons. It is clearly the case that many organizations are unable to master change, but ultimately such organizations will not be the ones to thrive.

You write that "different aspects of technology and humanity will advance at different rates, with the slower holding back the faster." I agree with the first part, but not the second. There's no question but that different parts of society evolve at different rates. We still have people pushing

plows with oxen, but the continued existence of preindustrial societies has not appreciably slowed down Intel and other companies from advancing microprocessor design.

In my view, the rigid cultural and religious factors that you eloquently describe end up being like stones in a stream. The water just flows around them. A good case in point is the current stem-cell controversy. Although I believe that banning therapeutic cloning represents an ignorant and destructive position, it has had the effect of accelerating workable approaches to converting one type of cell into another. Every cell has the complete genetic code, and we are beginning to understand the protein-signaling factors that control differentiation. The holy grail of tissue engineering will be to directly convert one cell into another by manipulating these signaling factors and thereby bypassing fetal tissue and egg cells altogether. We're not that far from being able to do this, and the current controversy has actually spurred these efforts. Ultimately, these will be superior approaches anyway because egg cells are hard to come by.

I think our differences here are rather subtle, and I agree strongly with your insight that "the derailing of progress is likely to be temporary and local, but no less real and potentially deadly for many."

On a different note, you ask, "Who gave us this. . .innate inclination for salvation." I agree that we're *evolved* rather than explicitly *designed* beings, so we can view this inclination to express "deep yearnings" as representative of our position as the cutting edge of evolution. We are that part of evolution that will lead the Universe to converting its endless masses of dumb matter into sublimely intelligent patterns of mass and energy. So we can view this inclination for transcendence in its evolutionary perspective as a useful survival tool in our ecological niche, a special niche for a species that is capable of modeling and extending its own capabilities.

Finally, I have to strongly endorse your conclusion that 'the sooner we master these technologies, the sooner we will conquer aging and death and all the evils that humankind has been heir to."

Max: I would like to respond to one point. You wrote:

"I do want to comment on your term 'the remaining human future' (being '25 years or 50 years or more'). I would rather consider the post-Singularity period to be one that is 'post biological' rather than 'post human.' In my view, the other side of the Singularity may properly be considered still human and still infused with (our better) human values. At least that is what we need to strive for."

I am sympathetic to what you are saying about the term "posthuman." It could carry connotations for some readers that it means disposing of all human values. Certainly no one could get that impression from what I have written on values, but the term itself may imply that. "Post-biological" is better in that sense, and I sometimes use that. Conceptually, I do think it is possible to be literally posthuman without yet being fully post-biological. For example, thorough genetic engineering, or a biological body supplemented by technological components, might be so divergent from the human genome that the person is no longer of the same species.

I wrote a paper on the human-species concept in light of where we are heading. I didn't get it into a form that quite satisfies me, but it did lead me to study the biologist's definitions of species concepts. (They cannot agree it seems!) I think it is reasonable by those definitions to talk of "posthuman." However, since the connotation may be undesirable, it may best be avoided. As you say, I would expect our post-Singularity selves to retain some of our human values (the better ones I hope). If we were to completely detach the human-species concept from its biological roots, we might still talk of post-Singularity humans, though I find that awkward. I did once try out the term "ultrahuman." That has the advantage of implying the retention of the best of humanity. However, I decided it sounded a bit too much like a superhero.

So, while I agree that "post-biological" works well for the most part, I'm still not quite settled on a preferred term. Any "post-" term is unsatisfying, but of course it's hard to create a positive term before we really know what forms we might take. I suspect that species concepts for ourselves may come to be fairly useless, post-Singularity.

Ray: I basically agree with what you're saying. Terminology is important, of course. For example, calling the multicellular bundles that can be used for creating stem cells "human embryos" has a lot of unfortunate implications. And recall that "nuclear magnetic resonance" was changed to "magnetic resonance imaging." When we figure out that strong magnetic fields have negative consequences, they'll probably have to change it again. We'll both have to work further on the terminology for the next stage in the evolution of our civilization.

Max: Ray, I want to thank you for inviting me into this engaging dialog. Such a vigorous yet well-mannered debate has, I think, helped both of us to detail our views further. Our shared premises and modestly differing conclusions have allowed us to tease out some implicit assumptions. As you concluded, our views amount to a rather small difference considered from the point of view of most people. The size of our divergence in estimations of time until that set of events we call the Singularity will seem large only once we are close to it. However, before then I expect our views to continue converging. What I find especially interesting is how we have reached very similar conclusions from quite different backgrounds. That someone with a formal background originally in economics and philosophy converges in thought with someone with a strong background in science and technology encourages me to favor E. O. Wilson's view of consilience. I look forward to continuing this dialog with you. It has been a pleasure and an honor.

Ray: Thanks, Max, for sharing your compelling thoughts. The feelings of satisfaction are mutual, and I look forward to continued convergence and consilience.

[1] http://www.kurzweilai.net/articles/art0092.html

Analysis One:

As noted, our initial observations are:

(1) $V = C1 \cdot W$

(2) $W = C2 \cdot \int_0^t V\,dt$

This gives us:

$$W = C1 \cdot C2 \cdot \int_0^t W\,dt$$

$$W = C1 \cdot C2 \cdot C3^{(C4\,t)}$$

$$V = C1 \cdot C2 \cdot C3^{(C4\,t)}$$

Analysis Two:

Let's factor in another exponential phenomenon, which is the growing resources for computation. Not only is each (constant cost) device getting more powerful as a function of W, but the resources deployed for computation are also growing exponentially.

We now have:

N: Expenditures for computation

$$V = C1 \cdot W$$

(as before)

$$N = C4^{(C5 \cdot t)}$$

(Expenditure for computation is growing at its own exponential rate)

$$W := C2 \cdot \int_0^t (N - V)\, dt$$

As before, world knowledge is accumulating, and the instantaneous increment is proportional to the amount of computation, which equals the resources deployed for computation (N) * the power of each (constant cost) device.

This gives us:

$$W = C1 \cdot C2 \cdot \left[\int_0^t \left(C4^{C5 \cdot t} \cdot W \right) dt \right]$$

$$W = C1 \cdot C2 \cdot \left[C3^{(C6 \cdot t)} \right]^{(C7 \cdot t)}$$

$$V = C1^2 \cdot C2 \cdot \left[C3^{(C6 \cdot t)} \right]^{(C7 \cdot t)}$$

Analysis Three:

Considering the data for actual calculating devices and computers during the twentieth century:

Let S = CPS/$1K: Calculations Per Second for $1,000

Twentieth century computing data matches:

$$S = 10^{\left[6.00 \cdot \left[\left(\frac{20.40}{6.00}\right)^{\left[\frac{(Year-1900)}{100}\right]}\right] - 11.00\right]}$$

We can determine the growth rate, G, over a period of time:

$$G = 10^{\left(\frac{\log(Sc) - \log(Sp)}{Yc - Yp}\right)}$$

Where Sc is CPS/$1K for current year, Sp is CPS/$1K for previous year, Yc is current year, and Yp is previous year

Human Brain = 100 Billion (10^{11}) neurons * 1000 (10^{3}) Connections/Neuron * 200 (2 * 10^{2}) Calculations Per Second Per Connection = 2 * 10^{16} Calculations Per Second

Human Race = 10 Billion (10^{10}) Human Brains = 2 * 10^{26} Calculations Per Second

We achieve one Human Brain capability (2 * 10^{16} cps) for $1,000 around the year 2023.

We achieve one Human Brain capability (2 * 10^{16} cps) for one cent around the year 2037.

We achieve one Human Race capability (2 * 10^{26} cps) for $1,000 around the year 2049.

Analysis Four:

(Note the following omits the constants)

$\frac{dW}{dt} = V$ giving $dW/c^{W} = dt$

This solves to

$W = \log\left(\frac{-1}{t}\right)$ $\qquad V = \frac{-1}{t}$

W and V rise very slowly when t<<0, which might be mistaken for an exponential around t = −1, and have a *singularity* at t = 0.

Analysis Five:

(Note the following omits the constants)

$V = W.W$

also $\frac{dW}{dt} = V + 1$ as before

This solves to $W = \tan(t)$ and $V = \tan(t)^2$, as before

Which has a singularity at $t = \frac{\pi}{2}$

Dangerous Futures

Will future technology – such as bioengineered pathogens, self-replicating nanobots, and supersmart robots – run amuck and accelerate out of control, perhaps threatening the human race? That's the concern of the pessimists, as stated by Bill Joy in an April 2000 Wired article. The optimists, such as Ray Kurzweil, believe technological progress is inevitable and can be controlled.

Are We Becoming an Endangered Species? Technology and Ethics in the Twenty First Century, A Panel Discussion at Washington National Cathedral

Ray Kurzweil

http://www.kurzweilai.net/articles/art0358.html

Ray Kurzweil addresses questions presented at Are We Becoming an Endangered Species? Technology and Ethics in the 21st Century, a conference on technology and ethics sponsored by Washington National Cathedral. Other panelists are Anne Foerst, Bill Joy and Bill McKibben.

Published on KurzweilAI.net November 19, 2001. Originally presented on November 19, 2001 at Washington National Cathedral. See the briefing paper, which contains questions posed to all panelists. Also see news item.

Bill McKibben, Ray Kurzweil, Judy Woodruff, Bill Joy, and Anne Foerst discuss the dangers of genetic engineering, nanotechnology and robotics.

Ray Kurzweil: Questions and Answers

Ray Kurzweil, how do you respond to Mr. Joy's concerns? Do scientific and technological advances pose a real threat to humanity, or do they promise to enhance life?

The answer is both, and we don't have to look further than today to see what I call the deeply intertwined promise and peril of technology.

Imagine going back in time, let's say a couple hundred years, and describing the dangers that lay ahead, perils such as weapons capable of destroying all mammalian life on Earth. People in the eighteenth century listening to this litany of dangers, assuming they believed you, would probably think it mad to take such risks.

And then you could go on and describe the actual suffering that lay ahead, for example 100 million people killed in two great twentieth-century world wars, made possible by technology, and so on. Suppose further that we provide these people circa eighteenth century a choice to relinquish these then future technologies, they just might choose to do so, particularly if we were to emphasize the painful side of the equation.

Our eighteenth century forbears, if provided with the visions of a reliable futurist of that day, and if given a choice, might very well have embraced the view of my fellow panelist Bill McKibben who says today that we "must now grapple squarely with the idea of a world that has enough wealth and enough technological capability, and should not pursue more."

Judy Woodruff interviews Ray Kurzweil at Washington National Cathedral.

Now I believe that implementing such a choice would require a Brave New World type of totalitarian government in which the government uses technology to ban the further development of technology, but let's put that perspective aside for a moment, and pursue this scenario further. What if our forefathers, and foremothers, had made such a decision? Would that have been so bad?

Well, for starters, most of us here today would not be here today, because life expectancy would have remained what it was back then, which was about 35 years of age. Furthermore, you would have been busy with the extraordinary toil and labor of everyday life with many hours required just to prepare the evening meal. The vast majority of humanity pursued lives that were labor-intensive, poverty-stricken, disease-ridden, and disaster-prone.

This basic equation has not changed. Technology has to a great extent liberated at least many of us from the enormous difficulty and fragility that characterized human life up until recent times. But there is still a great deal of affliction and distress that needs to be conquered, and that indeed can be overcome by technological advances that are close at hand. We are on the verge of multiple revolutions in biotechnology — genomics, proteomics, rational drug design, therapeutic cloning of cells, tissues, and organs, and others — that will save tens of millions of lives and alleviate enormous suffering. Ultimately, nanotechnology will provide the ability to create any physical product. Combined with other emerging technologies, we have the potential to largely eliminate poverty which also causes enormous misery.

And yes, as Bill Joy, and others, including myself, have pointed out, these same technologies can be applied in destructive ways, and invariably they will be. However, we have to be mindful of

the fact that our defensive technologies and protective measures will evolve along with the offensive potentials. If we take the future dangers such as Bill and others have described, and imagine them foisted on today's unprepared world, then it does sound like we're doomed. But that's not the delicate balance that we're facing. The defense will evolve along with the offense. And I don't agree with Bill that defense is necessarily weaker than offense. The reality is more complex.

We do have one contemporary example from which we can take a measure of comfort. Bill Joy talks about the dangers of self-replication, and we do have today a new form of fully nonbiological self-replicating entity that didn't exist just a few decades ago: the computer virus. When this form of destructive intruder first appeared, strong concerns were voiced that as they became more sophisticated, software pathogens had the potential to overwhelm, even destroy, the computer network medium they live in. Yet the immune system that has evolved in response to this challenge has been largely effective. The injury is but a small fraction of the benefit we receive from computer technology. That would not be the case if one imagines today's sophisticated software viruses foisted on the unprepared world of six or seven years ago.

One might counter that computer viruses do not have the lethal potential of biological viruses or self-replicating nanotechnology. Although true, this only strengthens my observation. The fact that computer viruses are usually not deadly to humans means that our response to the danger is that much less intense. Conversely, when it comes to self-replicating entities that are potentially lethal, our response on all levels will be vastly more serious.

Having said all this, I do have a specific proposal that I would like to share, which I will introduce a little later in our discussion.

Mr. Kurzweil, given humanity's track record with chemical and biological weapons, are we not guaranteed that terrorists and/or malevolent governments will abuse GNR (Genetic, Nanotechnology, Robotics) technologies? If so, how do we address this problem without an outright ban on the technologies?

Yes, these technologies will be abused. However, an outright ban, in my view, would be destructive, morally indefensible, and in any event would not address the dangers.

Nanotechnology, for example, is not a specific well-defined field. It is simply the inevitable end-result of the trend toward miniaturization which permeates virtually all technology. We've all seen pervasive miniaturization in our lifetimes. Technology in all forms — electronic, mechanical, biological, and others — is shrinking, currently at a rate of 5.6 per linear dimension per decade. The inescapable result will be nanotechnology.

With regard to more intelligent computers and software, it's an inescapable economic imperative affecting every company from large firms like Sun and Microsoft to small emerging companies.

With regard to biotechnology, are we going to tell the many millions of cancer sufferers around the world that although we are on the verge of new treatments that may save their lives, we're nonetheless canceling all of this research.

Banning these new technologies would condemn not just millions, but billions of people to the anguish of disease and poverty that we would otherwise be able to alleviate. And attempting to ban these technologies won't even eliminate the danger because it will only push these technologies underground where development would continue unimpeded by ethics and regulation.

We often go through three stages in examining the impact of future technology: awe and wonderment at its potential to overcome age-old problems, then a sense of dread at a new set of grave dangers that accompany these new technologies, followed by the realization that the only viable and responsible path is to set a careful course that can realize the promise while managing the peril.

The only viable approach is a combination of strong ethical standards, technology-enhanced law enforcement, and, most importantly, the development of both technical safeguards and technological immune systems to combat specific dangers.

And along those lines, I have a specific proposal. I do believe that we need to increase the priority of developing defensive technologies, not just for the vulnerabilities that society has identified since September 11, which are manifold, but the new ones attendant to the emerging technologies we're discussing this evening. We spend hundreds of billions of dollars a year on defense, and the danger from abuse of GNR technologies should be a primary target of these expenditures. Specifically, I am proposing that we set up a major program to be administered by the National Science Foundation and the National Institutes of Health. This new program would have a budget equaling the current budget for NSF and NIH. It would be devoted to developing defensive strategies, technologies, and ethical standards addressed at specific identified dangers associated with the new technologies funded by the conventional NSF and NIH budgets. There are other things we need to do as well, but this would be a practical way of significantly increasing the priority of addressing the dangers of emerging technologies.

If humans are going to play God, perhaps we should look at who is in the game. Mr. Kurzweil, isn't it true that both the technological and scientific fields lack broad participation by women, lower socioeconomic classes and sexual and ethnic minorities? If so, shouldn't we be concerned about the missing voices? What impact does the narrowly defined demographic have on technology and science?

I think it would be great to have more women in science, and it would lead to better decision making at all levels. To take an extreme example of the impact of not having sufficient participation by women, the Taliban have had no women in decision-making roles, and look at the quality of their decision-making.

To return to our own society, there are more women today in computer science, life sciences, and other scientific fields compared to 20 years ago, but clearly more progress is needed. With regard to ethnic groups such as Afro-Americans, the progress has been even less satisfactory, and I agree that addressing this is an urgent problem.

However, the real issue goes beyond direct participation in science and engineering. It has been said that war is too important to leave to the generals. It is also the case that science and engineering is too important to leave to the scientists and engineers. The advancement of technology from both the public and private sectors has a profound impact on every facet of our lives, from the nature of sexuality to the meaning of life and death.

To the extent that technology is shaped by market forces, then we all play a role as consumers. To the extent that science policy is shaped by government, then the political process is influential. But in order for everyone to play a role in playing God, there does need to be a meaningful dialog. And this in turn requires building bridges from the often incomprehensible world of scientific terminology to the everyday world that the educated lay public can understand.

Your work, Anne (Foerst), is unique and important in this regard, in that you've been building a bridge from the world of theology to the world of artificial intelligence, two seemingly disparate but surprisingly related fields. And Judy (Woodruff), journalism is certainly critical in that most people get their understanding of science and technology from the news.

We have many grave vulnerabilities in our society already. We can make a long list of exposures, and the press has been quite active in reporting on these since September 11. This does, incidentally, represent somewhat of a dilemma. On the one hand, reporting on these dangers is the way in which a democratic society generates the political will to address problems. On the other hand, if I were a terrorist, I would be reading the New York Times, and watching CNN, to get ideas and suggestions on the myriad ways in which society is susceptible to attack.

However, with regard to the GNR dangers, I believe this dilemma is somewhat alleviated because the dangers are further in the future. Now is the ideal time to be debating these emerging risks. It is also the right time to begin laying the scientific groundwork to develop the actual safeguards and defenses. We urgently need to increase the priority of this effort. That's why I've proposed a specific action item that for every dollar we spend on new technologies that can improve our lives, we spend another dollar to protect ourselves from the downsides of those same technologies.

How do you view the intrinsic worth of a "post-biological" world?

We've heard some discussion this evening on the dangers of ethnic and gender chauvinism. Along these lines, I would argue against human chauvinism and even biological chauvinism. On the other hand, I also feel that we need to revere and protect our biological heritage. And I do believe that these two positions are not incompatible.

We are in the early stages of a deep merger between the biological and nonbiological world. We already have replacement parts and augmentations for most of the organs and systems in our bodies. There is a broad variety of neural implants already in use. I have a deaf friend who I can now speak to on the telephone because of his cochlear implant. And he plans to have it upgraded to a new version that will provide a resolution of over a thousand frequencies that may restore

his ability to appreciate music. There are Parkinson's patients who have had their ability to move restored through an implant that replaces the biological cells destroyed by that disease.

By 2030, this merger of biological and nonbiological intelligence will be in high gear, and there will be many ways in which the two forms of intelligence work intimately together. So it won't be possible to come into a room and say, humans on the left, and machines on the right. There just won't be a clear distinction.

Since we're in a beautiful house of worship, let me relate this impending biological — nonbiological merger to a view of spiritual values.

I regard the freeing of the human mind from its severe physical limitations of scope and duration as the necessary next step in evolution. Evolution, in my view, represents the purpose of life. That is, the purpose of life — and of our lives — is to evolve.

What does it mean to evolve? Evolution moves toward greater complexity, greater elegance, greater knowledge, greater intelligence, greater beauty, greater creativity, and more of other abstract and subtle attributes such as love. And God has been called all these things, only without any limitation: all knowing, unbounded intelligence, infinite beauty, unlimited creativity, infinite love, and so on. Of course, even the accelerating growth of evolution never quite achieves an infinite level, but as it explodes exponentially, it certainly moves rapidly in that direction. So evolution moves inexorably closer to our conception of God, albeit never quite reaching this ideal. Thus the freeing of our thinking from the severe limitations of its biological form may be regarded as an essential spiritual quest.

One of the ways in which this universe of evolving patterns of matter and energy that we live in expresses its transcendent nature is in the exponential growth of the spiritual values we attribute in abundance to God: knowledge, intelligence, creativity, beauty, and love.

A Dialogue with the New York Times on the Technological Implications of the September 11 Disaster

Ray Kurzweil

http://www.kurzweilai.net/articles/art0327.html

In preparation for the New York Times article "In the Next Chapter, Is Technology an Ally?", Ray Kurzweil engaged in a conversation with computer scientist Peter Neumann, science fiction author Bruce Sterling, law professor Lawrence Lessig, retired engineer Severo Ornstein, and cryptographer Whitfield Diffie, addressing questions of how technology and innovation will be shaped by the tragic events of September 11, 2001.

Published on KurzweilAI.net September 27, 2001. Originally written September 22-23, 2001.

Question One from New York Times

Where do you think last week's terrorist attacks will take us in terms of technological innovation? What place is there for private entrepreneurship? Or will this see a resurgence of innovation at Government labs like Los Alamos, LLNL and Sandia, which have been losing their biggest talent to the private sector in recent years?

Ray Kurzweil's Response to Question One

The "September 11 tragedy" will accelerate a profound trend already well under way from centralized technologies to distributed ones, and from the real world to the virtual world. Centralized technologies involve an aggregation of resources such as people (e.g., cities, buildings), energy (e.g., nuclear power plants, liquid natural gas and oil tankers, energy pipelines), transportation (e.g., airplanes, trains), and other resources. Centralized technologies are subject to disruption and disaster. They also tend to be inefficient, wasteful, and harmful to the environment.

Distributed technologies, on the other hand, tend to be flexible, efficient, and relatively benign in their environment effects. The quintessential distributed technology is the Internet. Despite concerns about viruses, these information-based pathogens are mere nuisances. The Internet is essentially indestructible. If any hub or channel goes down, the information simply routes around it. The Internet is remarkably resilient, a quality that continues to grow with its continued exponential growth.

In the immediate aftermath of this crisis, we already see a dramatic movement away from meetings and conferences in the real world to those in the virtual world, including web-based meetings, Internet-based videoconferencing, and other contemporary examples of virtual communication. Meeting in this way is obviously safer, and ultimately more convenient. Despite the recent collapse of market value in telecommunications, bandwidth nonetheless continues to

expand exponentially which will continue to improve the resolution and sense of realism in the virtual world. We'll see a great deal of innovation to overcome many of the current limitations.

By the end of this decade, we'll have images written directly to our retinas from our eyeglasses and contact lenses, very high bandwidth wireless connection to the Internet available at all times, and the electronics for all this woven into our clothing. So we'll have ubiquitous, always-available, full-immersion, visual-auditory, shared virtual reality environments where we will be able to gather together for purposes ranging from business conferences to intimate encounters. The understandable fear from this tragedy is not going to go away, and will accelerate this trend.

In energy, we need to move rapidly toward the opposite end of the spectrum of contemporary energy sources, away from the extremely concentrated energy installations we now depend on. Many of today's energy technologies represent vulnerabilities far more grave than what we have just witnessed. As an example of a trend in the right direction, one company is pioneering fuel cells that are microscopic in size using MEMS (Micro Electronic Mechanical Systems) technology. They are manufactured like electronic chips but they are actually batteries with an energy to size ratio vastly exceeding conventional technology. Ultimately technology along these lines could power everything from our cell phones to our cars and homes. This type of technology would not be subject to disaster or disruption.

As these technologies develop, our need for aggregating people in large buildings and cities will diminish and people will spread out, living where they want, and gathering together in virtual reality. This is not a matter of simply giving in to Terrorist-generated fear, but rather a positive trend that will ultimately enhance the quality of life.

Question Two from New York Times

In the wake of the attacks, I've been hearing people say that we have been blinded by technology, that when we paint scenarios of terror, they tend to be laden with technology. Yet what these people did was in fact quite low tech. Have we become too smitten with a high-tech view of the world?

Ray Kurzweil's Response to Question Two

The terrorists clearly obtained remarkable leverage with their low tech weaponry. But the leverage came from the technology they commandeered (i.e., jet planes, jet fuel, huge buildings). Little attention has been paid to the security of our technology-based society, and the "opportunities" for such destructive leverage, unfortunately, are manifold. I share Peter Neumann's dilemma in wondering how much these leverage points should be publicly discussed. On the one hand, the only way to generate the political support to take the necessary security steps is through a public debate. On the other hand, no one wants to give the wrong people the wrong ideas.

I think we need to look at technology from the broadest perspective of its deeply intertwined promise and peril. Because of its inherently accelerating nature, most technology development in human history has occurred in the last two centuries. Compare life today to that of 200 years ago.

Life expectancy then was less than half of what it is today, and everyday life was extremely labor intensive (preparing the evening meal took much of the day), disease and poverty filled, and disaster-prone. Technology has liberated much of humanity from this precarious and painful situation. On the other hand, the "peril" side of technology provides concentrated power to create suffering on unprecedented scales. We've already seen this in the twentieth century. Hitler's trains and Stalin's tanks were applications of technology. Technology empowers both our creative and destructive impulses.

The issue is acute because we are not dealing with a static situation. Technology is accelerating (according to my models we're doubling the rate of paradigm shift rate, i.e., the rate of technology progress, every decade). There are already means available to cause outrages at a far greater scale than the tragedy we've just witnessed, as the discussion on terrorism using weapons of mass destructive makes clear. As we go forward, the same technology that will save millions of lives (and ease the enormous suffering) from cancer and other painful diseases will also potentially provide the means for a terrorist to create a bioengineered pathogen, which would again raise the stakes.

It is worth pointing out that we have not yet even dealt with the scenario that we witnessed on September 11. A terrorist can still take a plastic knife (i.e., I'm not talking about picnic plastic ware, but rather knives as effective as metal ones) through airport metal detectors. We have essentially no security with regard to private planes. The full list of vulnerabilities in our open society is very extensive.

We urgently need to identify these exposures and risks and develop defenses while also greatly augmenting public preparedness for different forms of terrorism, particularly those involving chemical and biological weapons and other weapons of mass destruction.

I began a conversation with Bill Joy in a Lake Tahoe Bar in October of 1998 on the intertwined promise and peril of twenty-first century technologies, a dialog which has continued in diverse venues. Although Bill and I are often paired as pessimist and optimist respectively, we actually agree on the reality of the dangers. September 11 is a wake up call, although I would say that we are still not taking seriously enough the diverse nature of the threats. I have been critical of Joy's apparent recommendation of relinquishment, and continue to be. Relinquishing broad areas of technology (such as nanotechnology) is neither feasible (not without relinquishing essentially all of technology) nor desirable. It would just drive these technologies underground where all the expertise would be left to the least responsible practitioners (i.e., the terrorists). However, I do support what I called "fine-grained relinquishment," which is avoiding specific capabilities and scenarios of particular danger (e.g., the Foresight Institute's call for ethical guidelines against creating entities that can self-replicate in natural environments). However, we also need to unleash the full power of our creativity on the defensive technologies. We also need to emphasize the relatively safer distributed technologies, such as distributed energy (e.g., microscopic sized fuel cells), and distributed communication (e.g., the Internet).

Question Three from New York Times

Larry Lessig says that the hard question is whether future innovation will be tailored to protect privacy as well as support legitimate state interests in surveillance and control.

Do you agree with him that we as a culture tend to think too crudely about technologies for surveillance? Where do you think the trade-offs should be?

And how, as Larry proposes, might the innovators develop technologies that both reserved important aspects of our freedom while responding to the real threats presented by the attacks."?

Ray Kurzweil's Response to Question Three

The nature of these terrorist attacks and the organization of the organization behind it puts civil liberties in general at odds with legitimate state interests in surveillance and control. The entire basis of our law enforcement system, and indeed much of our thinking about security, is based on an assumption that people are motivated to preserve their own lives and well being. That is the logic behind all of our strategies from law enforcement on the local level to "mutual assured destruction" on the world stage. But a foe that values the destruction of both its enemy and itself is not amenable to this line of attack.

So consider one very practical and current issue. The FBI identifies a likely terrorist cell and arrests the participants, even though they have not yet committed a crime and there may be insufficient evidence to convict them of a crime. Attorney General Ashcroft has proposed legislation that would allow the Government to continue to hold these individuals. The New York Times in its lead editorial today (September 23) objects to this and calls this a "troubling provision." So, according to the logic of this editorial, the Government should release these people because they have not yet committed a crime, and should re-arrest them only after they have committed a crime. Of course, by that time these terrorists will be dead along with a large number of their victims. How can the Government possibly break up a vast network of decentralized cells of suicide terrorists if they have to wait for each one to commit a crime?

I say this as someone who is generally very supportive of civil liberties. Indeed, one can point out that curtailing civil liberties in this way is exactly the aim of the terrorists, who despise our freedoms and our pluralistic society. Yet I do not see the prospect of any "magic bullet" innovation that would essentially change this equation.

The encryption trap door may be considered a technical "innovation" that the Government has been proposing in an attempt to balance legitimate individual needs for privacy with the government's need for surveillance. I have been supportive of this approach. Along with this type of technology we also need the requisite political innovation to provide for effective oversight by both the judicial and legislative branches of the executive branch's use of these trap doors to avoid the potential for abuse of power. The secretive nature of this opponent and its lack of respect for human life including its own will deeply test the foundation of our democratic traditions.

News item In the Next Chapter, Is Technology an Ally?

One Half of An Argument

Ray Kurzweil

http://www.kurzweilai.net/articles/art0236.html

A counterpoint to Jaron Lanier's dystopian visions of runaway technological cataclysm in "One Half of a Manifesto."

Excerpts from Jaron Lanier's "One Half of a Manifesto" are published on KurzweilAI.net at http://www.kurzweilai.net/articles/art0232.html along with his postscript on Ray Kurzweil at http://www.kurzweilai.net/articles/art0233.html.

Published on KurzweilAI.net July 31, 2001. Originally published on http://www.edge.org September 2000.

In Jaron Lanier's Postscript, which he wrote after he and I spoke in succession at a Vanguard event, Lanier points out that we agree on many things, which indeed we do. So I'll start in that vein as well. First of all, I share the world's esteem for Jaron's pioneering work in virtual reality, including his innovative contemporary work on the "Teleimmersion" initiative, and, of course, in coining the term "virtual reality." I probably have higher regard for virtual reality than Jaron does, but that comes back to our distinct views of the future.

As an aside I'm not entirely happy with the phrase "virtual reality" as it implies that it's not a real place to be. I consider a telephone conversation as being together in auditory virtual reality, yet we regard these to be real conversations. I have a similar problem with the term "artificial intelligence."

And as a pioneer in what I believe will become a transforming concept in human communication, I know that Jaron shares with me an underlying enthusiasm for the contributions that computer and related communications technologies can have on the quality of life. That is the other half of his manifesto. I appreciate Jaron pointing this out. It's not entirely clear sometimes, for example, that Bill Joy has another half to his manifesto.

And I agree with at least one of Jaron's six objections to what he calls "Cybernetic Totalism." In objection #3, he takes issues with those who maintain "that subjective experience either doesn't exist, or is unimportant because it is some sort of ambient or peripheral effect." The reason that some people feel this way is precisely because subjective experience cannot be scientifically measured. Although we can measure certain correlates of subjective experience (e.g., correlating certain patterns of objectively measurable neurological activity with objectively verifiable reports of certain subjective experiences), we cannot penetrate to the core of subjective experience through objective measurement. It's the difference between the concept of "objectivity," which is the basis of science, and "subjectivity," which is essentially a synonym for consciousness. There is no device or system we can postulate that could definitively detect subjectivity associated with an entity, at least no such device that does not have philosophical assumptions built into it.

So I accept that Jaron Lanier has subjective experiences, and I can even imagine (and empathize with!) his feelings of frustration at the dictums of "cybernetic totalists" such as myself (not that I accept this characterization) as he wrote his half manifesto. Like Jaron, I even accept the subjective experience of those who maintain that there is no such thing as subjective experience. Of course, most people do accept that other people are conscious, but this shared human assumption breaks down as we go outside of human experience, e.g., the debates regarding animal rights (which have everything to do with whether animals are conscious or just quasi-machines that operate by "instinct"), as well as the debates regarding the notion that a nonbiological entity could conceivably be conscious.

Consider that we are unable to truly experience the subjective experiences of others. We hear their reports about their experiences, and we may even feel empathy in response to the behavior that results from their internal states. We are, however, only exposed to the behavior of others and, therefore, can only imagine their subjective experience. So one can construct a perfectly consistent, and scientific, worldview that omits the existence of consciousness. And because there is fundamentally no scientific way to measure the consciousness or subjective experience of another entity, some observers come to the conclusion that it's just an illusion.

My own view is that precisely because we cannot resolve issues of consciousness entirely through objective measurement and analysis, i.e., science, there is a critical role for philosophy, which we sometimes call religion. I would agree with Jaron that consciousness is the most important ontological question. After all, if we truly imagine a world in which there is no subjective experience, i.e., a world in which there is swirling stuff but no conscious entity to experience it, then that world may as well not exist. In some philosophical traditions (i.e., some interpretations of quantum mechanics, some schools of Buddhist thought), that is exactly how such a world is regarded.

I like Jaron's term "circle of empathy," which makes it clear that the circle of reality that I consider to be "me" is not clear-cut. One's circle of empathy is certainly not simply our body, as we have limited identification with, say, our toes, and even less with the contents of our large intestines. Even with regard to our brains, we are aware of only a small portion of what goes on in our brains, and often consider thoughts and dreams that suddenly intrude on our awareness to have come from some foreign place. We do often include loved ones who may be physically disparate within our circle of empathy. Thus the aspect of the Universe that I consider to be "myself" is not at all clear cut, and some philosophies do emphasize the extent to which there is inherently no such boundary.

Having stated a few ways in which Jaron and I agree with each other's perspective, I will say that his "Half of a Manifesto" mischaracterizes many of the views he objects to. Certainly that's true with regard to his characterization of my own thesis. In particular, he appears to have only picked up on half of what I said in Atlanta, because the other half addresses at least some of the issues he raises. Moreover, many of Jaron's arguments aren't really arguments at all, but an amalgamation of mentally filed anecdotes and engineering frustrations. The fact that *Time* magazine got a prediction wrong in 1966, as Jaron reports, is not a compelling argument that all discussions of trends are misguided. Nor is the fact that dinosaurs did not continue to increase in size indefinitely a demonstration that every trend quickly dies out. The size of dinosaurs is

irrelevant; a larger size may or may not impart an advantage, whereas an increase in the price-performance and/or bandwidth of a technology clearly does impart an advantage. It would be hard to make the case that a technology with a lower price-performance had inherent advantages, whereas it is certainly possible that a smaller and therefore more agile animal may have advantages.

Jaron Lanier has what my colleague Lucas Hendrich calls the "engineer's pessimism." Often an engineer or scientist who is so immersed in the difficulties of a contemporary challenge fails to appreciate the ultimate long-term implications of their own work, and, in particular, the larger field of work that they operate in. Consider the biochemists in 1985 who were skeptical of the announcement of the goal of transcribing the entire genome in a mere 15 years. These scientists had just spent an entire year transcribing a mere one ten-thousandth of the genome, so even with reasonable anticipated advances, it seemed to them like it would be hundreds of years, if not longer, before the entire genome could be sequenced. Or consider the skepticism expressed in the mid 1980s that the Internet would ever be a significant phenomenon, given that it included only tens of thousands of nodes. The fact that the number of nodes was doubling every year and there were, therefore, likely to be tens of millions of nodes ten years later was not appreciated by those who struggled with "state of the art" technology in 1985, which permitted adding only a few thousand nodes throughout the world in a year.

In his "Postscript regarding Ray Kurzweil," Jaron asks the rhetorical question "about Ray's exponential theory of history...[is he] stacking the deck by choosing points that fit the curves he wants to find?" I can assure Jaron that the more points we add to the dozens of exponential graphs I presented to him and the rest of the audience in Atlanta, the clearer the exponential trends become. Does he really imagine that there is some circa 1901 calculating device that has better price-performance than our circa 2001 devices? Or even a 1995 device that is competitive with a 2001 device? In fact what we do see as more points (representing specific devices) are collected is a cascade of "S-curves," in which each S-curve represents some specific technological paradigm. Each S-curve (which looks like an "S" in which the top portion is stretched out to the right) starts out with gradual and then extreme exponential growth, subsequently leveling off as the potential of that paradigm is exhausted. But what turns each S-curve into an ongoing exponential is the shift to another paradigm, and thus to another S-curve, i.e., innovation. The pressure to explore and discover a new paradigm increases as the limits of each current paradigm becomes apparent.

When it became impossible to shrink vacuum tubes any further and maintain the requisite vacuum, transistors came along, which are not merely small vacuum tubes. We've been through five paradigms in computing in this past century (electromechanical calculators, relay-based computers, vacuum-tube-based computing, discrete transistors, and then integrated circuits, on which Moore's law is based). As the limits of flat integrated circuits are now within sight (one to one and a half decades away), there are already dozens of projects underway to pioneer the sixth paradigm of computing, which is computing in three dimensions, several of which have demonstrated small-scale working systems.

It is specifically the processing and movement of information that is growing exponentially. So one reason that an area such as transportation is resting at the top of an S-curve is that many if

not most of the purposes of transportation have been satisfied by exponentially growing communication technologies. My own organization has colleagues in different parts of the country, and most of our needs that in times past would have required a person or a package to be transported can be met through the increasingly viable virtual meetings made possible by a panoply of communication technologies, some of which Jaron is himself working to advance. Having said that, I do believe we will see new paradigms in transportation. However, with increasingly realistic, high-resolution full-immersion forms of virtual reality continuing to emerge, our needs to be together will increasingly be met through computation and communication.

Jaron's concept of "lock-in" is not the primary obstacle to advancing transportation. If the existence of a complex support system necessarily caused lock-in, then why don't we see lock-in preventing ongoing expansion of every aspect of the Internet? After all, the Internet certainly requires an enormous and complex infrastructure. The primary reason that transportation is under little pressure for a paradigm-shift is that the underlying need for transportation has been increasingly met through communication technologies that are expanding exponentially.

One of Jaron's primary themes is to distinguish between quantitative and qualitative trends, saying in essence that perhaps certain brute force capabilities such as memory capacity, processor speed, and communications bandwidths are expanding exponentially, but the qualitative aspects are not. And toward this end, Jaron complains of a multiplicity of software frustrations (many, incidentally, having to do with Windows) that plague both users and, in particular, software developers like himself.

This is the hardware versus software challenge, and it is an important one. Jaron does not mention at all my primary thesis having to do with the software of intelligence. Jaron characterizes my position and that of other so-called "cybernetic totalists" to be that we'll just figure it out in some unspecified way, what he refers to as a software "Deus ex Machina." I have a specific and detailed scenario to achieve the software of intelligence, which concerns the reverse engineering of the human brain, an undertaking that is much further along than most people realize. I'll return to this in a moment, but first I would like to address some other basic misconceptions about the so-called lack of progress in software.

Jaron calls software inherently "unwieldy" and "brittle" and writes at great length on a variety of frustrations that he encounters in the world of software. He writes that "getting computers to perform specific tasks of significant complexity in a reliable but modifiable way, without crashes or security breaches, is essentially impossible." I certainly don't want to put myself in the position of defending all software (any more than I would care to characterize all people as wonderful). But it's not the case that complex software is necessarily brittle and prone to catastrophic breakdown. There are many examples of complex mission critical software that operates with very little if any breakdowns, for example the sophisticated software that controls an increasing fraction of airplane landings, or software that monitors patients in critical care facilities. I am not aware of any airplane crashes that have been caused by automated landing software; the same, however, cannot be said for human reliability.

Jaron says that "Computer user interfaces tend to respond more slowly to user interface events, such as a key press, than they did fifteen years ago...What's gone wrong?" To this I would invite Jaron to try using an old computer today. Even we put aside the difficulty of setting one up today (which is a different issue), Jaron has forgotten just how unresponsive, unwieldy, and limited they were. Try getting some real work done to today's standards with a fifteen year-old personal computer. It's simply not true to say that the old software was better in any qualitative or quantitative sense. If you believe that, then go use them.

Although it's always possible to find poor quality design, the primary reason for user interface response delays is user demand for more sophisticated functionality. If users were willing to freeze the functionality of their software, then the ongoing exponential growth of computing speed and memory would quickly eliminate software response delays. But they're not. So functionality always stays on the edge of what's feasible (personally, I'm waiting for my Teleimmersion upgrade to my videoconferencing software).

This romancing of software from years or decades ago is comparable to people's idyllic view of life hundreds of years ago, when we were unencumbered with the frustrations of machines. Life was unencumbered, perhaps, but it was also short (e.g., life expectancy less than half of today's), labor-intensive (e.g., just preparing the evening meal took many hours of hard labor), poverty-filled, disease and disaster prone.

With regard to the price-performance of software, the comparisons in virtually every area are dramatic. For example, in 1985 $5,000 bought you a speech recognition software package that provided a 1,000 word vocabulary, no continuous speech capability, required three hours of training, and had relatively poor accuracy. Today, for only $50, you can purchase a speech recognition software package with a 100,000 word vocabulary, continuous speech, that requires only five minutes of training, has dramatically improved accuracy, natural language understanding ability (for editing commands and other purposes), and many other features.

How about software development itself? I've been developing software myself for forty years, so I have some perspective on this. It's clear that the growth in productivity of software development has a lower exponent, but it is nonetheless growing exponentially. The development tools, class libraries, and support systems available today are dramatically more effective than those of decades ago. I have today small teams of just three or four people who achieve objectives in a few months that are comparable to what a team of a dozen or more people could accomplish in a year or more 25 years ago. I estimate the doubling time of software productivity to be approximately six years, which is slower than the doubling time for processor price-performance, which is approximately one year today. However, software productivity is nonetheless growing exponentially.

The most important point to be made here is that there is a specific game plan for achieving human-level intelligence in a machine. I agree that achieving the requisite hardware capacity is a necessary but not sufficient condition. As I mentioned above, we have a resource for understanding how to program the methods of human intelligence given hardware that is up to the task, and that resource is the human brain itself.

Here again, if you speak to some of the neurobiologists who are diligently creating detailed mathematical models of the hundreds of types of neurons found in the brain, or who are modeling the patterns of connections found in different regions, you will in at least a few cases encounter the same sort of engineer's/scientist's myopia that results from being immersed in the specifics of one aspect of a large challenge. However, having tracked the progress being made in accumulating all of the (yes, exponentially increasing) knowledge about the human brain and its algorithms, I believe that it is a conservative scenario to expect that within thirty years we will have detailed models of the several hundred information processing organs we collectively call the human brain.

For example, Lloyd Watts has successfully synthesized (that is, assembled and integrated) the detailed models of neurons and interconnections in more than a dozen regions of the brain having to do with auditory processing. He has a detailed model of the information transformations that take place in these regions, and how this information is encoded, and has implemented these models in software. The performance of Watt's software matches the intricacies that have been revealed in subtle experiments on human hearing and auditory discrimination. Most interestingly, using Watts' models as the front-end in speech recognition has demonstrated the ability to pick out one speaker against a backdrop of background sounds, an impressive feat that humans are capable of, and that up until Watts' work, had not been feasible in automated speech recognition systems.

The brain is not one big neural net. It consists of hundreds of regions, each of which is organized differently, with different types of neurons, different types of signaling, and different patterns of interconnections. By and large, the algorithms are not the sequential, logical methods that are commonly used in digital computing. The brain tends to use self-organizing, chaotic, holographic (i.e., information not in one place but distributed throughout a region), massively parallel, and digital-controlled-analog methods. However, we have demonstrated in a wide range of projects the ability to understand these methods, and to extract them from the rapidly escalating knowledge of the brain and its organization.

The speed, cost effectiveness, and bandwidth of human brain scanning is also growing exponentially, doubling every year. Our knowledge of human neuron models is also rapidly growing. The size of neuron clusters that we have successfully recreated in terms of functional equivalence is also scaling up exponentially.

I am not saying that this process of reverse engineering the human brain is the only route to "strong" AI. It is, however, a critical source of knowledge that is feeding into our overall research activities where these methods are integrated with other approaches.

Also, it is not the case that the complexity of software, and therefore its "brittleness" needs to scale up dramatically in order to emulate the human brain, even when we get to emulating its full functionality. My own area of technical interest is pattern recognition, and the methods that we typically use are self-organizing methods such as neural nets, Markov models, and genetic algorithms. When set up in the right way, these methods can often display subtle and complex behaviors that are not predictable by the designer putting them into practice. I'm not saying that such self-organizing methods are an easy short cut to creating complex and intelligent behavior,

but they do represent one important way in which the complexity of a system can be increased without the brittleness of explicitly programmed logical systems.

Consider that the brain itself is created from a genome with only 23 million bytes of useful information (that's what's left of the 800 million byte genome when you eliminate all the redundancies, e.g., the sequence "ALU" which is repeated hundreds of thousands of times). 23 million bytes is not that much information (it's less than Microsoft Word). How is it, then, that the human brain with its 100 trillion connections can result from a genome that is so small? I have estimated that just the interconnection data alone to characterize the human brain is a million times greater than the information in the genome.

The answer is that the genome specifies a set of processes, each of which utilizes chaotic methods (i.e., initial randomness, then self-organization) to increase the amount of information represented. It is known, for example, that the wiring of the interconnections follows a plan that includes a great deal of randomness. As the individual person encounters her environment, the connections and the neurotransmitter level patterns self-organize to better represent the world, but the initial design is specified by a program that is not extreme in its complexity.

It is not my position that we will program human intelligence link by link as in some huge CYC-like expert system. Nor is it the case that we will simply set up a huge genetic (i.e., evolutionary) algorithm and have intelligence at human levels automatically evolve itself. Jaron worries correctly that any such approach would inevitably get stuck in some local minima. He also interestingly points out how biological evolution "missed the wheel." Actually, that's not entirely accurate. There are small wheel-like structures at the protein level, although it's true that their primary function is not for vehicle transportation. Wheels are not very useful, of course, without roads. However, biological evolution did create a species that created wheels (and roads), so it did succeed in creating a lot of wheels, albeit indirectly (but there's nothing wrong with indirect methods, we use them in engineering all the time).

With regard to creating human levels of intelligence in our machines, we will integrate the insights and models gained from reverse engineering the human brain, which will involve hundreds of regions, each with different methods, many of which do involve self-organizing paradigms at different levels. The feasibility of this reverse engineering project and of implementing the revealed methods has already been clearly demonstrated. I don't have room in this response to describe the methodology and status of brain reverse engineering in detail, but I will point out that the concept is not necessarily limited to neuromorphic modeling of each neuron. We can model substantial neural clusters by implementing parallel algorithms that are functionally equivalent. This often results in substantially reduced computational requirements, which has been shown by Lloyd Watts and Carver Mead.

Jaron writes that "if there ever was a complex, chaotic phenomenon, we are it." I agree with that, but don't see this as an obstacle. My own area of interest is chaotic computing, which is how we do pattern recognition, which in turn is the heart of human intelligence. Chaos is part of the process of pattern recognition, it drives the process, and there is no reason that we cannot harness these methods in our machines just as they are utilized in our brains.

Jaron writes that "evolution has evolved, introducing sex, for instance, but evolution has never found a way to be any speed but very slow." But he is ignoring the essential nature of an evolutionary process, which is that it accelerates because each stage introduces more powerful methods for creating the next stage. Biological evolution started out extremely slow, and the first halting steps took billions of years. The design of the principal body plans was faster, requiring only tens of millions of years. The process of biological evolution has accelerated, with each stage faster than the stage before it. Later key steps, such as the emergence of homo sapiens, took only hundreds of thousands of years. Human technology, which is evolution continued indirectly (created by a species created by evolution), continued this acceleration. The first steps took tens of thousands of years, outpacing biological evolution, and has accelerated from there. The World Wide Web emerged in only a few years, distinctly faster than, say, the Cambrian explosion.

Jaron complains that "surprisingly few of the most essential algorithms have overheads that scale at a merely linear rate." Without taking up several pages to analyze this statement in detail, I will point out that the brain does what it does in its own real-time, using interneuronal connections (where most of our thinking takes place) that operate at least ten million times slower than contemporary electronic circuits. We can observe the brain's massively parallel methods in detail, ultimately scan and understand all of its tens of trillions of connections, and replicate its methods. As I've mentioned, we're well down that path.

To correct a few of Jaron's statements regarding (my) time frames, it's not my position that the "singularity" will "arrive a quarter of the way into the new century" or that a "new criticality" will be "achieved in the about the year 2020." Just so that the record is straight, my view is that we will have the requisite hardware capability to emulate the human brain in a $1,000 of a computation (which won't be organized in the rectangular forms we see today such as notebooks and palmtops, but rather embedded in our environment) by 2020. The software will take longer, to around 2030. The "singularity" has divergent definitions, but for our purposes here we can consider this to be a time when nonbiological forms of intelligence dominate purely biological forms, albeit being derivative of them. This takes us beyond 2030, to perhaps 2040 or 2050.

Jaron calls this an "immanent doom" and "an eschatological cataclysm," as if it were clear on its face that such a development were undesirable. I view these developments as simply the continuation of the evolutionary process and neither utopian nor dystopian. It's true, on the one hand, that nanotechnology and strong AI, and particularly the two together, have the potential to solve age-old problems of poverty and human suffering, not to mention clean up the messes we're creating today with some of our more primitive technologies. On the other hand, there will be profound new problems and dangers that will emerge as well. I have always considered technology to be a double-edged sword. It amplifies both our creative and destructive natures, and we don't have to look further than today to see that.

However, on balance, I view the progression of evolution as a good thing, indeed as a spiritual direction. What we see in evolution is a progression toward greater intelligence, greater creativity, greater beauty, greater subtlety (i.e., the emergence of entities with emotion such as the ability to love, therefore greater love). And "God" has been described as an ideal of an infinite level of these same attributes. Evolution, even in its exponential growth, never reaches

infinite levels, but it's moving rapidly in that direction. So we could say that this evolutionary process is moving in a spiritual direction.

However, the story of the twenty-first century has not yet been written. So it's not my view that any particular story is inevitable, only that evolution, which has been inherently accelerating since the dawn of biological evolution, will continue its exponential pace.

Jaron writes that "the whole enterprise of Artificial Intelligence is based on an intellectual mistake." Until such time that computers at least match human intelligence in every dimension, it will always remain possible for skeptics to say the glass is half empty. Every new achievement of AI can be dismissed by pointing out yet other goals have not yet been accomplished. Indeed, this is the frustration of the AI practitioner, that once an AI goal is achieved, it is no longer considered AI and becomes just a useful technique. AI is inherently the set of problems we have not yet solved.

Yet machines are indeed growing in intelligence, and the range of tasks that machines can accomplish that previously required intelligent human attention is rapidly growing. There are hundreds of examples of narrow AI today (e.g., computers evaluating electrocardiograms and blood cell images, making medical diagnoses, guiding cruise missiles, making financial investment decisions, not to mention intelligently routing emails and cell phone connections), and the domains are becoming broader. Until such time that the entire range of human intellectual capability is emulated, it will always be possible to minimize what machines are capable of doing.

I will point out that once we have achieved complete models of human intelligence, machines will be capable of combining the flexible, subtle, human levels of pattern recognition with the natural advantages of machine intelligence. For example, machines can instantly share knowledge, whereas we don't have quick downloading ports on our interconnection and neurotransmitter concentration level patterns. Machines are much faster (as I mentioned contemporary electronics is already ten million times faster than the electrochemical information processing used in our brains) and have much more prodigious and accurate memories.

Jaron refers to the annual "Turing test" that Loebner runs, and maintains that "we have caused the Turing test to be passed." These are misconceptions. I used to be on the prize committee of this contest until a political conflict caused most of the prize committee members to quit. Be that as it may, this contest is not really a Turing test, as we're not yet at that stage. It's a "narrow Turing test" which deals with domain-specific dialogues, not unrestricted dialog as Turing envisioned it. With regard to the Turing test as Turing described it, it is generally accepted that this has not yet happened.

Returning to Jaron's nice phrase "circle of empathy," he writes that his "personal choice is to not place computers inside the circle." But would he put neurons inside that circle? We've already shown that a neuron or even a substantial cluster of neurons can be emulated in great detail and accuracy by computers. So where on that slippery slope does Jaron find a stable footing? As Rodney Brooks says in his September 25, 2000 commentary on Jaron's Half of a Manifesto, Jaron "turns out to be a closet Searlean." He just assumes that a computer cannot be as subtle —

or as conscious — as the hundreds of neural regions we call the human brain. Like Searle, Jaron just assumes his conclusion. (For a more complete discussion of Searle and his theories, see my essay "Locked in his Chinese Room, Response to John Searle" in the forthcoming book Are We Spiritual Machines?: Ray Kurzweil vs. the Critics of Strong AI, Discovery Institute Press, 2001. This entire book is posted on http://www.kurzweilai.net/meme/memelist.html?m=19).

Near the end of Jaron's essay, he worries about the "terrifying" possibility that through these technologies the rich may obtain certain opportunities that the rest of humankind does not have access to. This, of course, would be nothing new, but I would point out that because of the ongoing exponential growth of price-performance, all of these technologies quickly become so inexpensive as to become almost free. Look at the extraordinary amount of high-quality information available at no cost on the web today which did not exist at all just a few years ago. And if one wants to point out that only a small fraction of the world today has Web access, keep in mind that the explosion of the Web is still in its infancy.

At the end of his Half of a Manifesto, Jaron writes that "the ideology of cybernetic totalist intellectuals [may] be amplified from novelty into a force that could cause suffering for millions of people." I don't believe this fearful conclusion follows from Jaron's half of an argument. The bottom line is that technology is power and this power is rapidly increasing. Technology may result in suffering or liberation, and we've certainly seen both in the twentieth century. I would argue that we've seen more of the latter, but nevertheless neither Jaron nor I wish to see the amplification of destructiveness that we have witnessed in the past one hundred years. As I mentioned above, the story of the twenty first century has not yet been written. I think Jaron would agree with me that our destiny is in our hands. However, I regard "our hands" to include our technology, which is properly part of the human-machine civilization.

Jaron Lanier's "One Half of a Manifesto" was originally published September 2000 at Edge.

Nanotechnology

Think small. The nanotechnology boom is beginning. Now how do we keep it under control?

Testimony of Ray Kurzweil on the Societal Implications of Nanotechnology

Ray Kurzweil

http://www.kurzweilai.net/articles/art0556.html

Despite calls to relinquish research in nanotechnology, we will have no choice but to confront the challenge of guiding nanotechnology in a constructive direction. Advances in nanotechnology and related advanced technologies are inevitable. Any broad attempt to relinquish nanotechnology will only push it underground, which would interfere with the benefits while actually making the dangers worse.

Published on KurzweilAI.net April 8, 2003. Testimony presented April 9, 2003 at the Committee on Science, U.S. House of Representatives Hearing to examine the societal implications of nanotechnology and consider H.R. 766, The Nanotechnology Research and Development Act of 2003.

Summary of Testimony:

The size of technology is itself inexorably shrinking. According to my models, both electronic and mechanical technologies are shrinking at a rate of 5.6 per linear dimension per decade. At this rate, most of technology will be "nanotechnology" by the 2020s.

We are immeasurably better off as a result of technology, but there is still a lot of suffering in the world to overcome. We have a moral imperative, therefore, to continue the pursuit of knowledge and advanced technologies, such as nanotechnology, that can continue to overcome human affliction. There is also an economic imperative to continue due to the pervasive acceleration of technology, including miniaturization, in the competitive economy.

Nanotechnology is not a separate field of study that we can simply relinquish. We will have no choice but to confront the challenge of guiding nanotechnology in a constructive direction. There are strategies we can deploy, but there will need to be continual development of defensive strategies.

We can take some level of comfort from our relative success in dealing with one new form of fully non-biological, self-replicating pathogen: the software virus.

The most immediate danger is not self-replicating nanotechnology, but rather self-replicating biotechnology. We need to place a much higher priority on developing vitally needed defensive technologies such as antiviral medications. Keep in mind that a bioterrorist does not need to put his "innovations" through the FDA.

Any broad attempt to relinquish nanotechnology will only push it underground, which would interfere with the benefits while actually making the dangers worse.

Existing regulations on the safety of foods, drugs, and other materials in the environment are sufficient to deal with the near-term applications of nanotechnology, such as nanoparticles.

Full Verbal Testimony:

Chairman Boehlert, distinguished members of the U.S. House of Representatives Committee on Science, and other distinguished guests, I appreciate this opportunity to respond to your questions and concerns on the vital issue of the societal implications of nanotechnology. Our rapidly growing ability to manipulate matter and energy at ever smaller scales promises to transform virtually every sector of society, including health and medicine, manufacturing, electronics and computers, energy, travel, and defense. There will be increasing overlap between nanotechnology and other technologies of increasing influence, such as biotechnology and artificial intelligence. As with any other technological transformation, we will be faced with deeply intertwined promise and peril.

In my brief verbal remarks, I only have time to summarize my conclusions on this complex subject, and I am providing the Committee with an expanded written response that attempts to explain the reasoning behind my views.

Eric Drexler's 1986 thesis developed the concept of building molecule-scale devices using molecular assemblers that would precisely guide chemical reactions. Without going through the history of the controversy surrounding feasibility, it is fair to say that the consensus today is that nano-assembly is indeed feasible, although the most dramatic capabilities are still a couple of decades away.

The concept of nanotechnology today has been expanded to include essentially any technology where the key features are measured in a modest number of nanometers (under 100 by some definitions). By this standard, contemporary electronics has already passed this threshold.

For the past two decades, I have studied technology trends, along with a team of researchers who have assisted me in gathering critical measures of technology in different areas, and I have been developing mathematical models of how technology evolves. Several conclusions from this study have a direct bearing on the issues before this hearing. Technologies, particularly those related to information, develop at an exponential pace, generally doubling in capability and price-performance every year. This observation includes the power of computation, communication – both wired and wireless, DNA sequencing, brain scanning, brain reverse engineering, and the size and scope of human knowledge in general. Of particular relevance to this hearing, the size of technology is itself inexorably shrinking. According to my models, both electronic and mechanical technologies are shrinking at a rate of 5.6 per linear dimension per decade. At this rate, most of technology will be "nanotechnology" by the 2020s.

The golden age of nanotechnology is, therefore, a couple of decades away. This era will bring us the ability to essentially convert software, i.e., information, directly into physical products. We will be able to produce virtually any product for pennies per pound. Computers will have greater computational capacity than the human brain, and we will be completing the reverse engineering

of the human brain to reveal the software design of human intelligence. We are already placing devices with narrow intelligence in our bodies for diagnostic and therapeutic purposes. With the advent of nanotechnology, we will be able to keep our bodies and brains in a healthy, optimal state indefinitely. We will have technologies to reverse environmental pollution. Nanotechnology and related advanced technologies of the 2020s will bring us the opportunity to overcome age-old problems, including pollution, poverty, disease, and aging.

We hear increasingly strident voices that object to the intermingling of the so-called natural world with the products of our technology. The increasing intimacy of our human lives with our technology is not a new story, and I would remind the committee that had it not been for the technological advances of the past two centuries, most of us here today would not be here today. Human life expectancy was 37 years in 1800. Most humans at that time lived lives dominated by poverty, intense labor, disease, and misfortune. We are immeasurably better off as a result of technology, but there is still a lot of suffering in the world to overcome. We have a moral imperative, therefore, to continue the pursuit of knowledge and of advanced technologies that can continue to overcome human affliction.

There is also an economic imperative to continue. Nanotechnology is not a single field of study that we can simply relinquish, as suggested by Bill Joy's essay, "Why the Future Doesn't Need Us." Nanotechnology is advancing on hundreds of fronts, and is an extremely diverse activity. We cannot relinquish its pursuit without essentially relinquishing all of technology, which would require a Brave New World totalitarian scenario, which is inconsistent with the values of our society.

Technology has always been a double-edged sword, and that is certainly true of nanotechnology. The same technology that promises to advance human health and wealth also has the potential for destructive applications. We can see that duality today in biotechnology. The same techniques that could save millions of lives from cancer and disease may also empower a bioterrorist to create a bioengineered pathogen.

A lot of attention has been paid to the problem of self-replicating nanotechnology entities that could essentially form a nonbiological cancer that would threaten the planet. I discuss in my written testimony steps we can take now and in the future to ameliorate these dangers. However, the primary point I would like to make is that we will have no choice but to confront the challenge of guiding nanotechnology in a constructive direction. Any broad attempt to relinquish nanotechnology will only push it underground, which would interfere with the benefits while actually making the dangers worse.

As a test case, we can take a small measure of comfort from how we have dealt with one recent technological challenge. There exists today a new form of fully nonbiological self-replicating entity that didn't exist just a few decades ago: the computer virus. When this form of destructive intruder first appeared, strong concerns were voiced that as they became more sophisticated, software pathogens had the potential to destroy the computer network medium they live in. Yet the "immune system" that has evolved in response to this challenge has been largely effective. Although destructive self-replicating software entities do cause damage from time to time, the injury is but a small fraction of the benefit we receive from the computers and

communication links that harbor them. No one would suggest we do away with computers, local area networks, and the Internet because of software viruses.

One might counter that computer viruses do not have the lethal potential of biological viruses or of destructive nanotechnology. This is not always the case: we rely on software to monitor patients in critical care units, to fly and land airplanes, to guide intelligent weapons in our current campaign in Iraq, and other "mission critical" tasks. To the extent that this is true, however, this observation only strengthens my argument. The fact that computer viruses are not usually deadly to humans only means that more people are willing to create and release them. It also means that our response to the danger is that much less intense. Conversely, when it comes to self-replicating entities that are potentially lethal on a large scale, our response on all levels will be vastly more serious, as we have seen since 9-11.

I would describe our response to software pathogens as effective and successful. Although they remain (and always will remain) a concern, the danger remains at a nuisance level. Keep in mind that this success is in an industry in which there is no regulation, and no certification for practitioners. This largely unregulated industry is also enormously productive. One could argue that it has contributed more to our technological and economic progress than any other enterprise in human history.

Some of the concerns that have been raised, such as Bill Joy's article, are effective because they paint a picture of future dangers as if they were released on today's unprepared world. The reality is that the sophistication and power of our defensive technologies and knowledge will grow along with the dangers.

The challenge most immediately in front of us is not self-replicating nanotechnology, but rather self-replicating biotechnology. The next two decades will be the golden age of biotechnology, whereas the comparable era for nanotechnology will follow in the 2020s and beyond. We are now in the early stages of a transforming technology based on the intersection of biology and information science. We are learning the "software" methods of life and disease processes. By reprogramming the information processes that lead to and encourage disease and aging, we will have the ability to overcome these afflictions. However, the same knowledge can also empower a terrorist to create a bioengineered pathogen.

As we compare the success we have had in controlling engineered software viruses to the coming challenge of controlling engineered biological viruses, we are struck with one salient difference. As I noted, the software industry is almost completely unregulated. The same is obviously not the case for biotechnology. A bioterrorist does not need to put his "innovations" through the FDA. However, we do require the scientists developing the defensive technologies to follow the existing regulations, which slow down the innovation process at every step. Moreover, it is impossible, under existing regulations and ethical standards, to test defenses to bioterrorist agents on humans. There is already extensive discussion to modify these regulations to allow for animal models and simulations to replace infeasible human trials. This will be necessary, but I believe we will need to go beyond these steps to accelerate the development of vitally needed defensive technologies.

With the human genome project, 3 to 5 percent of the budgets were devoted to the ethical, legal, and social implications (ELSI) of the technology. A similar commitment for nanotechnology would be appropriate and constructive.

Near-term applications of nanotechnology are far more limited in their benefits as well as more benign in their potential dangers. These include developments in the materials area involving the addition of particles with multi-nanometer features to plastics, textiles, and other products. These have perhaps the greatest potential in the area of pharmaceutical development by allowing new strategies for highly targeted drugs that perform their intended function and reach the appropriate tissues, while minimizing side effects. This development is not qualitatively different than what we have been doing for decades in that many new materials involve constituent particles that are novel and of a similar physical scale. The emerging nanoparticle technology provides more precise control, but the idea of introducing new nonbiological materials into the environment is hardly a new phenomenon. We cannot say a priori that all nanoengineered particles are safe, nor would it be appropriate to deem them necessarily unsafe. Environmental tests thus far have not shown reasons for undue concern, and it is my view that existing regulations on the safety of foods, drugs, and other materials in the environment are sufficient to deal with these near-term applications.

The voices that are expressing concern about nanotechnology are the same voices that have expressed undue levels of concern about genetically modified organisms. As with nanoparticles, GMO's are neither inherently safe nor unsafe, and reasonable levels of regulation for safety are appropriate. However, none of the dire warnings about GMO's have come to pass. Already, African nations, such as Zambia and Zimbabwe, have rejected vitally needed food aid under pressure from European anti-GMO activists. The reflexive anti-technology stance that has been reflected in the GMO controversy will not be helpful in balancing the benefits and risks of nanoparticle technology.

In summary, I believe that existing regulatory mechanisms are sufficient to handle near-term applications of nanotechnology. As for the long term, we need to appreciate that a myriad of nanoscale technologies are inevitable. The current examinations and dialogues on achieving the promise while ameliorating the peril are appropriate and will deserve sharply increased attention as we get closer to realizing these revolutionary technologies.

Written Testimony

I am pleased to provide a more detailed written response to the issues raised by the committee. In this written portion of my response, I address the following issues:

- **Models of Technology Trends**: A discussion of why nanotechnology and related advanced technologies are inevitable. The underlying technologies are deeply integrated into our society and are advancing on many diverse fronts.
- **A Small Sample of Examples of True Nanotechnology**: a few of the implications of nanotechnology two to three decades from now.

- **The Economic Imperatives of the Law of Accelerating Returns**: the exponential advance of technology, including the accelerating miniaturization of technology, is driven by economic imperative, and, in turn, has a pervasive impact on the economy.
- **The Deeply Intertwined Promise and Peril of Nanotechnology and Related Advanced Technologies**: Technology is inherently a doubled-edged sword, and we will need to adopt strategies to encourage the benefits while ameliorating the risks. Relinquishing broad areas of technology, as has been proposed, is not feasible and attempts to do so will only drive technology development underground, which will exacerbate the dangers.

Models of Technology Trends

A diverse technology such as nanotechnology progresses on many fronts and is comprised of hundreds of small steps forward, each benign in itself. An examination of these trends shows that technology in which the key features are measured in a small number of nanometers is inevitable. I hereby provide some examples of my study of technology trends.

The motivation for this study came from my interest in inventing. As an inventor in the 1970s, I came to realize that my inventions needed to make sense in terms of the enabling technologies and market forces that would exist when the invention was introduced, which would represent a very different world than when it was conceived. I began to develop models of how distinct technologies – electronics, communications, computer processors, memory, magnetic storage, and the size of technology – developed and how these changes rippled through markets and ultimately our social institutions. I realized that most inventions fail not because they never work, but because their timing is wrong. Inventing is a lot like surfing, you have to anticipate and catch the wave at just the right moment.

In the 1980s, my interest in technology trends and implications took on a life of its own, and I began to use my models of technology trends to project and anticipate the technologies of future times, such as the year 2000, 2010, 2020, and beyond. This enabled me to invent with the capabilities of the future. In the late 1980s, I wrote my first book, *The Age of Intelligent Machines*, which ended with the specter of machine intelligence becoming indistinguishable from its human progenitors. This book included hundreds of predictions about the 1990s and early 2000 years, and my track record of prediction has held up well.

During the 1990s I gathered empirical data on the apparent acceleration of all information-related technologies and sought to refine the mathematical models underlying these observations. In *The Age of Spiritual Machines* (ASM), which I wrote in 1998, I introduced refined models of technology, and a theory I called "the law of accelerating returns," which explained why technology evolves in an exponential fashion.

The Intuitive Linear View versus the Historical Exponential View

The future is widely misunderstood. Our forebears expected the future to be pretty much like their present, which had been pretty much like their past. Although exponential trends did exist a thousand years ago, they were at that very early stage where an exponential trend is so flat and so

slow that it looks like no trend at all. So their lack of expectations was largely fulfilled. Today, in accordance with the common wisdom, everyone expects continuous technological progress and the social repercussions that follow. But the future will nonetheless be far more surprising than most observers realize because few have truly internalized the implications of the fact that the rate of change itself is accelerating.

Most long-range forecasts of technical feasibility in future time periods dramatically underestimate the power of future developments because they are based on what I call the "intuitive linear" view of history rather than the "historical exponential view." To express this another way, it is not the case that we will experience a hundred years of progress in the twenty-first century; rather we will witness on the order of twenty thousand years of progress (at *today's* rate of progress, that is).

When people think of a future period, they intuitively assume that the current rate of progress will continue for future periods. Even for those who have been around long enough to experience how the pace increases over time, an unexamined intuition nonetheless provides the impression that progress changes at the rate that we have experienced recently. From the mathematician's perspective, a primary reason for this is that an exponential curve approximates a straight line when viewed for a brief duration. It is typical, therefore, that even sophisticated commentators, when considering the future, extrapolate the current pace of change over the next 10 years or 100 years to determine their expectations. This is why I call this way of looking at the future the "intuitive linear" view.

But a serious assessment of the history of technology shows that technological change is exponential. In exponential growth, we find that a key measurement such as computational power is multiplied by a constant factor for each unit of time (e.g., doubling every year) rather than just being added to incrementally. Exponential growth is a feature of any evolutionary process, of which technology is a primary example. One can examine the data in different ways, on different time scales, and for a wide variety of technologies ranging from electronic to biological, as well as social implications ranging from the size of the economy to human life span, and the acceleration of progress and growth applies. Indeed, we find not just simple exponential growth, but "double" exponential growth, meaning that the rate of exponential growth is itself growing exponentially. These observations do not rely merely on an assumption of the continuation of Moore's law (i.e., the exponential shrinking of transistor sizes on an integrated circuit), but is based on a rich model of diverse technological processes. What it clearly shows is that technology, particularly the pace of technological change, advances (at least) exponentially, not linearly, and has been doing so since the advent of technology, indeed since the advent of evolution on Earth.

Many scientists and engineers have what my colleague Lucas Hendrich calls "engineer's pessimism." Often an engineer or scientist who is so immersed in the difficulties and intricate details of a contemporary challenge fails to appreciate the ultimate long-term implications of their own work, and, in particular, the larger field of work that they operate in. Consider the biochemists in 1985 who were skeptical of the announcement of the goal of transcribing the entire genome in a mere 15 years. These scientists had just spent an entire year transcribing a mere one ten-thousandth of the genome, so even with reasonable anticipated advances, it seemed

to them like it would be hundreds of years, if not longer, before the entire genome could be sequenced. Or consider the skepticism expressed in the mid 1980s that the Internet would ever be a significant phenomenon, given that it included only tens of thousands of nodes. The fact that the number of nodes was doubling every year and there were, therefore, likely to be tens of millions of nodes ten years later was not appreciated by those who struggled with "state of the art" technology in 1985, which permitted adding only a few thousand nodes throughout the world in a year.

I emphasize this point because it is the most important failure that would-be prognosticators make in considering future trends. The vast majority of technology forecasts and forecasters ignore altogether this "historical exponential view" of technological progress. Indeed, almost everyone I meet has a linear view of the future. That is why people tend to overestimate what can be achieved in the short term (because we tend to leave out necessary details), but underestimate what can be achieved in the long term (because the exponential growth is ignored).

The Law of Accelerating Returns

The ongoing acceleration of technology is the implication and inevitable result of what I call the "law of accelerating returns," which describes the acceleration of the pace and the exponential growth of the products of an evolutionary process. This includes technology, particularly information-bearing technologies, such as computation. More specifically, the law of accelerating returns states the following:

- Evolution applies positive feedback in that the more capable methods resulting from one stage of evolutionary progress are used to create the next stage. As a result, the rate of progress of an evolutionary process increases exponentially over time. Over time, the "order" of the information embedded in the evolutionary process (i.e., the measure of how well the information fits a purpose, which in evolution is survival) increases.
- A correlate of the above observation is that the "returns" of an evolutionary process (e.g., the speed, cost-effectiveness, or overall "power" of a process) increase exponentially over time.
- In another positive feedback loop, as a particular evolutionary process (e.g., computation) becomes more effective (e.g., cost effective), greater resources are deployed towards the further progress of that process. This results in a second level of exponential growth (i.e., the rate of exponential growth itself grows exponentially).
- Biological evolution is one such evolutionary process.
- Technological evolution is another such evolutionary process. Indeed, the emergence of the first technology-creating species resulted in the new evolutionary process of technology. Therefore, technological evolution is an outgrowth of–and a continuation of–biological evolution.
- A specific paradigm (a method or approach to solving a problem, e.g., shrinking transistors on an integrated circuit as an approach to making more powerful computers) provides exponential growth until the method exhausts its potential. When this happens, a paradigm shift (a fundamental change in the approach) occurs, which enables exponential growth to continue.

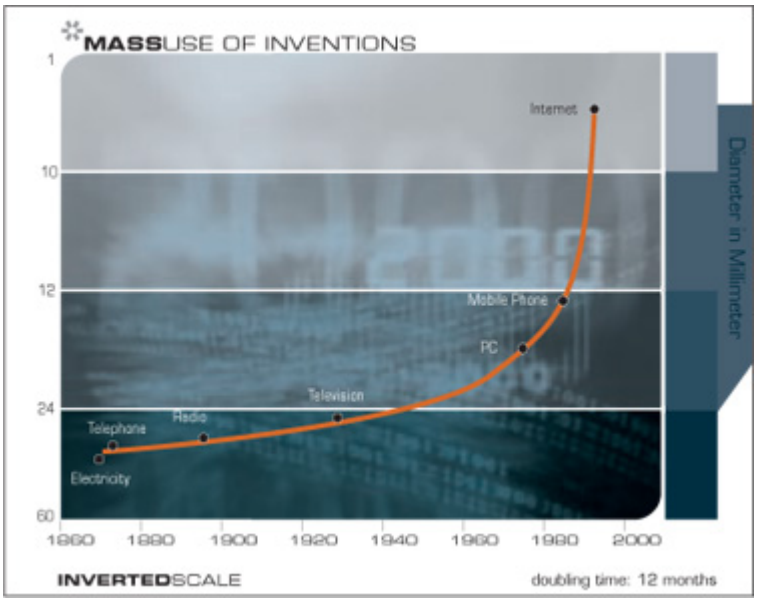
- Each paradigm follows an "S-curve," which consists of slow growth (the early phase of exponential growth), followed by rapid growth (the late, explosive phase of exponential growth), followed by a leveling off as the particular paradigm matures.
- During this third or maturing phase in the life cycle of a paradigm, pressure builds for the next paradigm shift.
- When the paradigm shift occurs, the process begins a new S-curve.
- Thus the acceleration of the overall evolutionary process proceeds as a sequence of S-curves, and the overall exponential growth consists of this cascade of S-curves.
- The resources underlying the exponential growth of an evolutionary process are relatively unbounded.
- One resource is the (ever-growing) order of the evolutionary process itself. Each stage of evolution provides more powerful tools for the next. In biological evolution, the advent of DNA allowed more powerful and faster evolutionary "experiments." Later, setting the "designs" of animal body plans during the Cambrian explosion allowed rapid evolutionary development of other body organs, such as the brain. Or to take a more recent example, the advent of computer-assisted design tools allows rapid development of the next generation of computers.
- The other required resource is the "chaos" of the environment in which the evolutionary process takes place and which provides the options for further diversity. In biological evolution, diversity enters the process in the form of mutations and ever-changing environmental conditions, including cosmological disasters (e.g., asteroids hitting the Earth). In technological evolution, human ingenuity combined with ever-changing market conditions keep the process of innovation going.

If we apply these principles at the highest level of evolution on Earth, the first step, the creation of cells, introduced the paradigm of biology. The subsequent emergence of DNA provided a digital method to record the results of evolutionary experiments. Then, the evolution of a species that combined rational thought with an opposable appendage (the thumb) caused a fundamental paradigm shift from biology to technology. The upcoming primary paradigm shift will be from biological thinking to a hybrid combining biological and nonbiological thinking. This hybrid will include "biologically inspired" processes resulting from the reverse engineering of biological brains.

If we examine the timing of these steps, we see that the process has continuously accelerated. The evolution of life forms required billions of years for the first steps (e.g., primitive cells); later on progress accelerated. During the Cambrian explosion, major paradigm shifts took only tens of millions of years. Later on, Humanoids developed over a period of millions of years, and Homo sapiens over a period of only hundreds of thousands of years.

With the advent of a technology-creating species, the exponential pace became too fast for evolution through DNA-guided protein synthesis and moved on to human-created technology. Technology goes beyond mere tool making; it is a process of creating ever more powerful technology using the tools from the previous round of innovation, and is, thereby, an evolutionary process. The first technological steps—sharp edges, fire, the wheel–took tens of thousands of years. For people living in this era, there was little noticeable technological change in even a thousand years. By 1000 AD, progress was much faster and a paradigm shift required

only a century or two. In the nineteenth century, we saw more technological change than in the nine centuries preceding it. Then in the first twenty years of the twentieth century, we saw more advancement than in all of the nineteenth century. Now, paradigm shifts occur in only a few years time. The World Wide Web did not exist in anything like its present form just a few years ago; it didn't exist at all a decade ago.



The paradigm shift rate (i.e., the overall rate of technical progress) is currently doubling (approximately) every decade; that is, paradigm shift times are halving every decade (and the rate of acceleration is itself growing exponentially). So, the technological progress in the twenty-first century will be equivalent to what would require (in the linear view) on the order of 200 centuries. In contrast, the twentieth century saw only about 20 years of progress (again at today's rate of progress) since we have been speeding up to current rates. So the twenty-first century will see about a thousand times greater technological change than its predecessor.

Moore's Law and Beyond

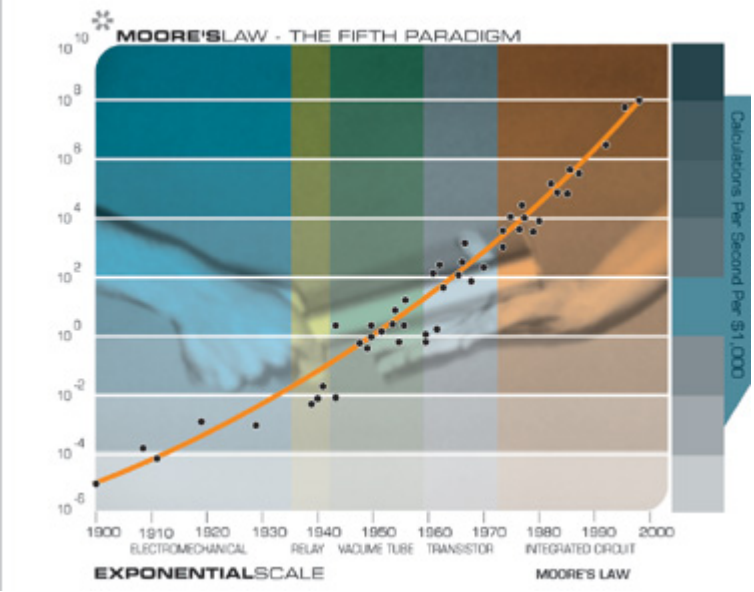
There is a wide range of technologies that are subject to the law of accelerating returns. The exponential trend that has gained the greatest public recognition has become known as "Moore's Law." Gordon Moore, one of the inventors of integrated circuits, and then Chairman of Intel, noted in the mid-1970s that we could squeeze twice as many transistors on an integrated circuit every 24 months. Given that the electrons have less distance to travel, the circuits also run twice as fast, providing an overall quadrupling of computational power.

However, the exponential growth of computing is much broader than Moore's Law.

If we plot the speed (in instructions per second) per $1000 (in constant dollars) of 49 famous calculators and computers spanning the entire twentieth century, we note that there were four completely different paradigms that provided exponential growth in the price-performance of computing before the integrated circuits were invented. Therefore, Moore's Law was not the

first, but the fifth paradigm to exponentially grow the power of computation. And it won't be the last. When Moore's Law reaches the end of its S-Curve, now expected before 2020, the exponential growth will continue with three-dimensional molecular computing, a prime example of the application of nanotechnology, which will constitute the sixth paradigm.

When I suggested in my book *The Age of Spiritual Machines*, published in 1999, that three-dimensional molecular computing, particularly an approach based on using carbon nanotubes, would become the dominant computing hardware technology in the teen years of this century, that was considered a radical notion. There has been so much progress in the past four years, with literally dozens of major milestones having been achieved, that this expectation is now a mainstream view.



Moore's Law Was Not the First, but the Fifth Paradigm to Provide Exponential Growth of Computing. Each time one paradigm runs out of steam, another picks up the pace

The exponential growth of computing is a marvelous quantitative example of the exponentially growing returns from an evolutionary process. We can express the exponential growth of computing in terms of an accelerating pace: it took 90 years to achieve the first MIPS (million instructions per second) per thousand dollars; now we add one MIPS per thousand dollars every day.

Moore's Law narrowly refers to the number of transistors on an integrated circuit of fixed size, and sometimes has been expressed even more narrowly in terms of transistor feature size. But rather than feature size (which is only one contributing factor), or even number of transistors, I think the most appropriate measure to track is computational speed per unit cost. This takes into account many levels of "cleverness" (i.e., innovation, which is to say, technological evolution). In addition to all of the innovation in integrated circuits, there are multiple layers of innovation in computer design, e.g., pipelining, parallel processing, instruction look-ahead, instruction and memory caching, and many others.

The human brain uses a very inefficient electrochemical digital-controlled analog computational process. The bulk of the calculations are done in the interneuronal connections at a speed of only about 200 calculations per second (in each connection), which is about ten million times slower than contemporary electronic circuits. But the brain gains its prodigious powers from its extremely parallel organization *in three dimensions*. There are many technologies in the wings that build circuitry in three dimensions. Nanotubes, an example of nanotechnology, which is already working in laboratories, build circuits from pentagonal arrays of carbon atoms. One cubic inch of nanotube circuitry would be a million times more powerful than the human brain. There are more than enough new computing technologies now being researched, including three-dimensional silicon chips, optical and silicon spin computing, crystalline computing, DNA computing, and quantum computing, to keep the law of accelerating returns as applied to computation going for a long time.

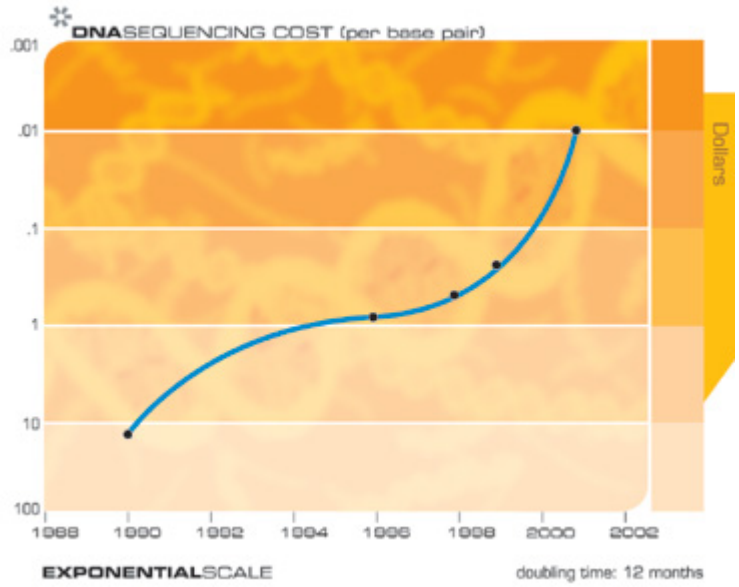
As I discussed above, it is important to distinguish between the "S" curve (an "S" stretched to the right, comprising very slow, virtually unnoticeable growth – followed by very rapid growth – followed by a flattening out as the process approaches an asymptote) that is characteristic of any specific technological paradigm and the continuing exponential growth that is characteristic of the ongoing evolutionary process of technology. Specific paradigms, such as Moore's Law, do ultimately reach levels at which exponential growth is no longer feasible. That is why Moore's Law is an S curve. But the growth of computation is an ongoing exponential (at least until we "saturate" the Universe with the intelligence of our human-machine civilization, but that will not be a limit in this coming century). In accordance with the law of accelerating returns, paradigm shift, also called innovation, turns the S curve of any specific paradigm into a continuing exponential. A new paradigm (e.g., three-dimensional circuits) takes over when the old paradigm approaches its natural limit, which has already happened at least four times in the history of computation. This difference also distinguishes the tool making of non-human species, in which the mastery of a tool-making (or using) skill by each animal is characterized by an abruptly ending S shaped learning curve, versus human-created technology, which has followed an exponential pattern of growth and acceleration since its inception.

DNA Sequencing, Memory, Communications, the Internet, and Miniaturization

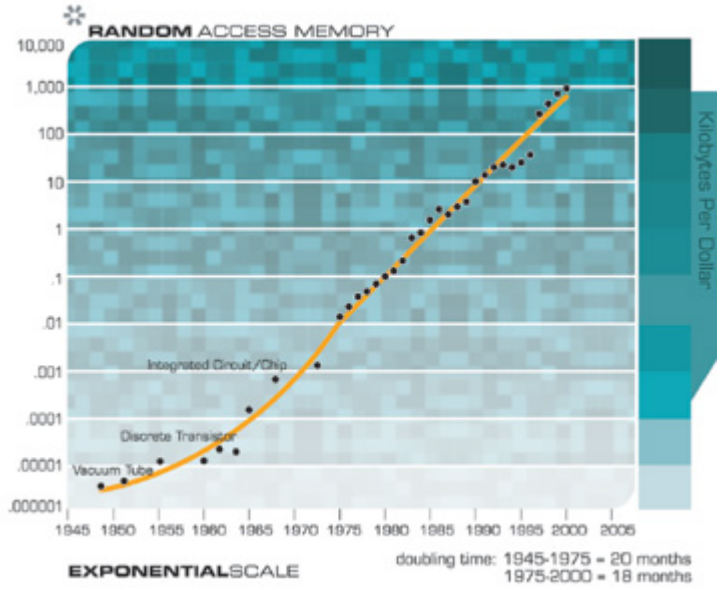
This "law of accelerating returns" applies to all of technology, indeed to any true evolutionary process, and can be measured with remarkable precision in information-based technologies. There are a great many examples of the exponential growth implied by the law of accelerating returns in technologies, as varied as DNA sequencing, communication speeds, brain scanning, electronics of all kinds, and even in the rapidly shrinking size of technology, which is directly relevant to the discussion at this hearing. The future nanotechnology age results not from the exponential explosion of computation alone, but rather from the interplay and myriad synergies that will result from manifold intertwined technological revolutions. Also, keep in mind that every point on the exponential growth curves underlying these panoply of technologies (see the graphs below) represents an intense human drama of innovation and competition. It is

remarkable therefore that these chaotic processes result in such smooth and predictable exponential trends.

As I noted above, when the human genome scan started fourteen years ago, critics pointed out that given the speed with which the genome could then be scanned, it would take thousands of years to finish the project. Yet the fifteen year project was nonetheless completed slightly ahead of schedule.

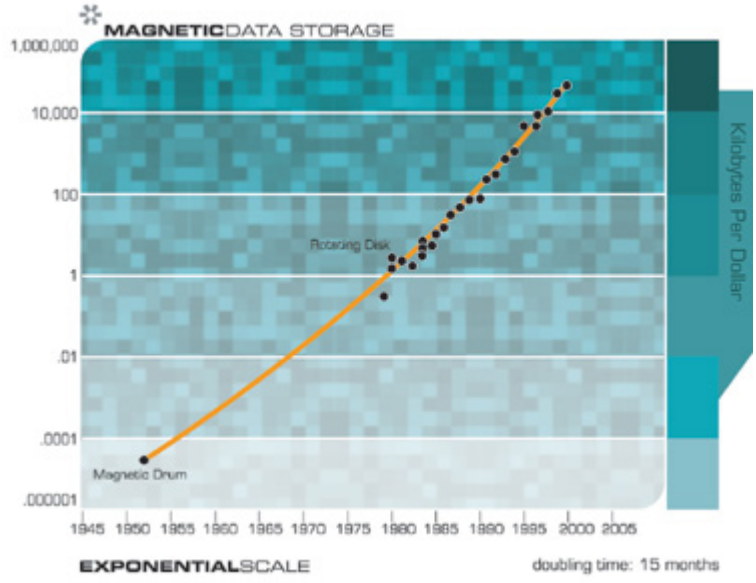


Of course, we expect to see exponential growth in electronic memories such as RAM.

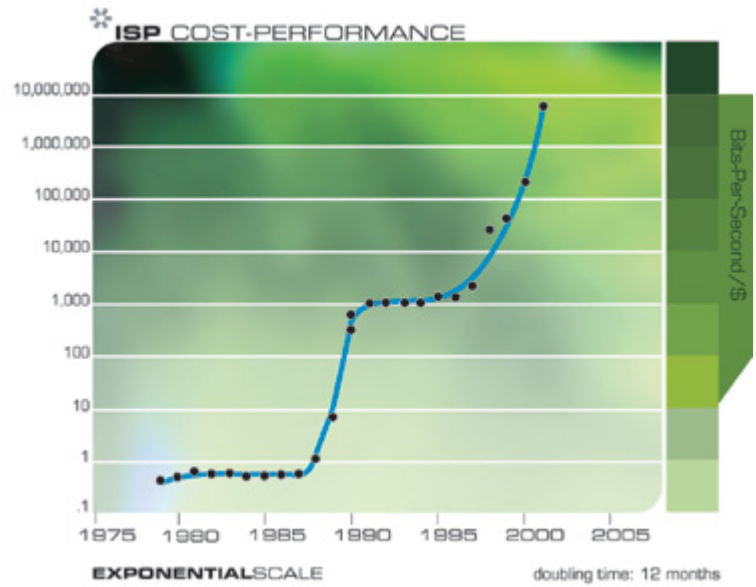


Notice How Exponential Growth Continued through Paradigm Shifts from Vacuum Tubes to Discrete Transistors to Integrated Circuits

However, growth in magnetic memory is not primarily a matter of Moore's law, but includes advances in mechanical and electromagnetic systems.



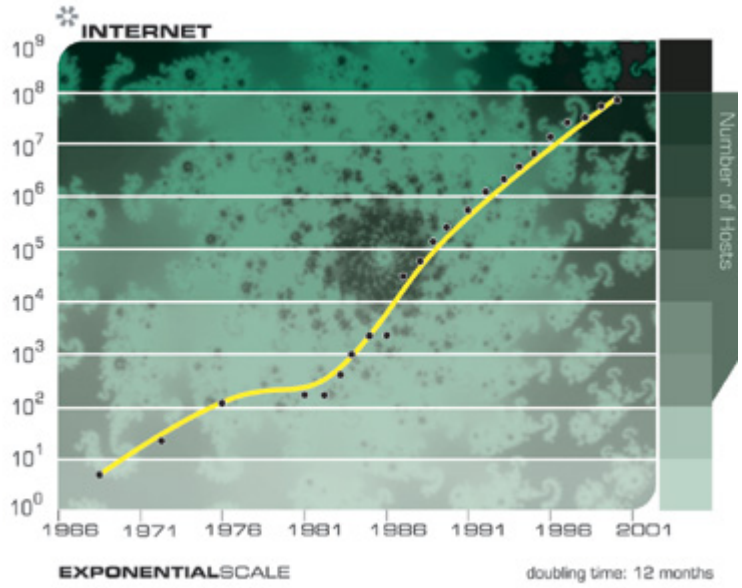
Exponential growth in communications technology has been even more explosive than in computation and is no less significant in its implications. Again, this progression involves far more than just shrinking transistors on an integrated circuit, but includes accelerating advances in fiber optics, optical switching, electromagnetic technologies, and others.

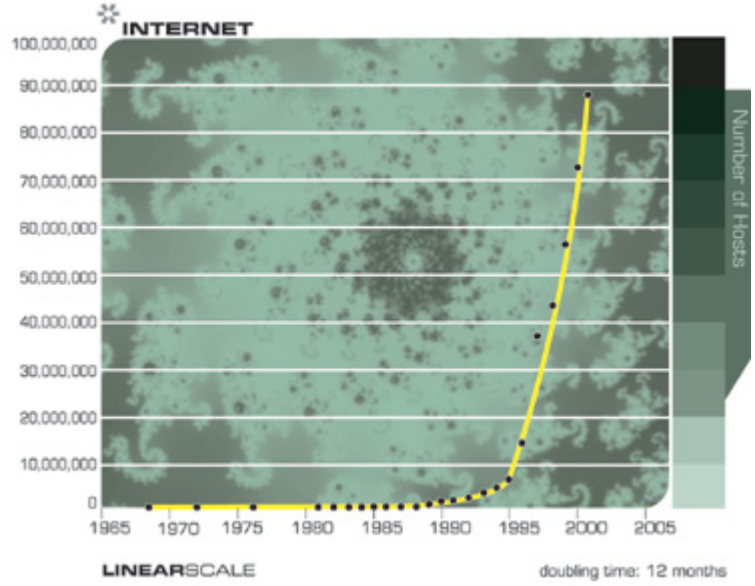


Notice Cascade of "S" Curves

Note that in the above chart we can actually see the progression of "S" curves: the acceleration fostered by a new paradigm, followed by a leveling off as the paradigm runs out of steam, followed by renewed acceleration through paradigm shift.

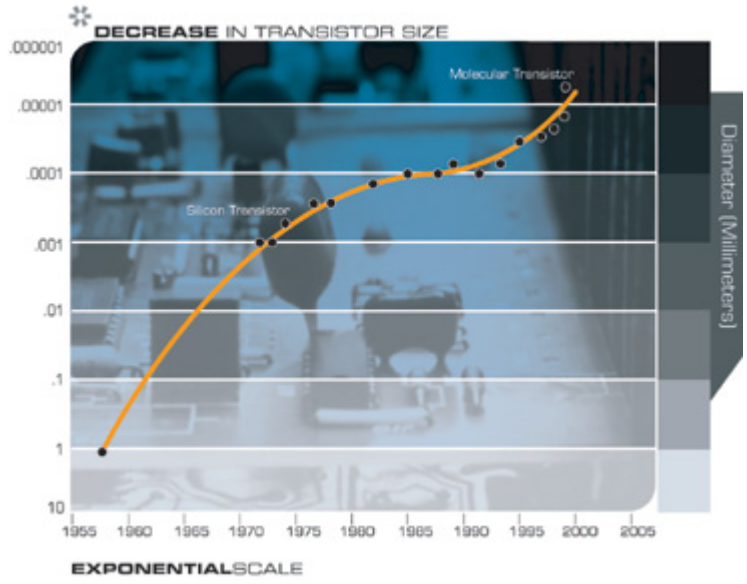
The following two charts show the overall growth of the Internet based on the number of hosts (server computers). These two charts plot the same data, but one is on an exponential axis and the other is linear. As I pointed out earlier, whereas technology progresses in the exponential domain, we experience it in the linear domain. So from the perspective of most observers, nothing was happening until the mid 1990s when seemingly out of nowhere, the World Wide Web and email exploded into view. But the emergence of the Internet into a worldwide phenomenon was readily predictable much earlier by examining the exponential trend data.

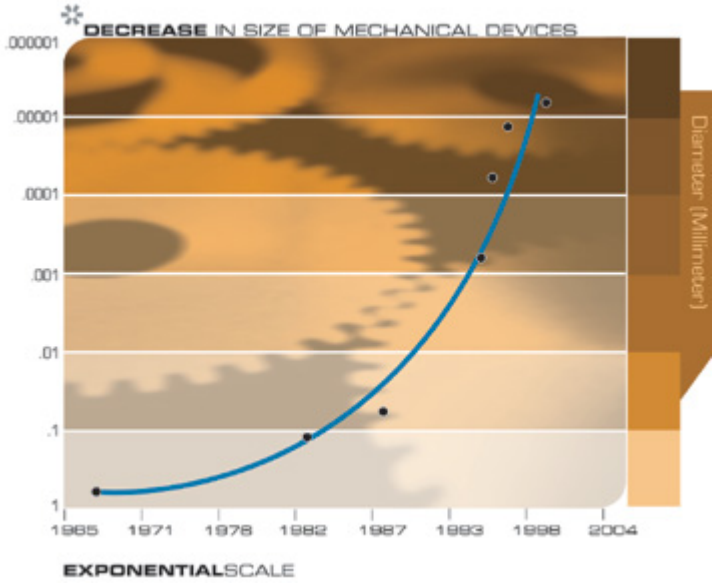




Notice how the explosion of the Internet appears to be a surprise from the Linear Chart, but was perfectly predictable from the Exponential Chart

The most relevant trend to this hearing, and one that will have profound implications for the twenty-first century is the pervasive trend towards making things smaller, i.e., miniaturization. The salient implementation sizes of a broad range of technologies, both electronic and mechanical, are shrinking, also at a double-exponential rate. At present, we are shrinking technology by a factor of approximately 5.6 per linear dimension per decade.





A Small Sample of Examples of True Nanotechnology

Ubiquitous nanotechnology is two to three decades away. A prime example of its application will be to deploy billions of "nanobots": small robots the size of human blood cells that can

travel inside the human bloodstream. This notion is not as futuristic as it may sound in that there have already been successful animal experiments using this concept. There are already four major conferences on "BioMEMS" (Biological Micro Electronic Mechanical Systems) covering devices in the human blood stream.

Consider several examples of nanobot technology, which, based on miniaturization and cost reduction trends, will be feasible within 30 years. In addition to scanning the human brain to facilitate human brain reverse engineering, these nanobots will be able to perform a broad variety of diagnostic and therapeutic functions inside the bloodstream and human body. Robert Freitas, for example, has designed robotic replacements for human blood cells that perform hundreds or thousands of times more effectively than their biological counterparts. With Freitas' "respirocytes," (robotic red blood cells), you could do an Olympic sprint for 15 minutes without taking a breath. His robotic macrophages will be far more effective than our white blood cells at combating pathogens. His DNA repair robot would be able to repair DNA transcription errors, and even implement needed DNA changes. Although Freitas' conceptual designs are two or three decades away, there has already been substantial progress on bloodstream-based devices. For example, one scientist has cured type I Diabetes in rats with a nanoengineered device that incorporates pancreatic Islet cells. The device has seven- nanometer pores that let insulin out, but block the antibodies which destroy these cells. There are many innovative projects of this type already under way.

Clearly, nanobot technology has profound military applications, and any expectation that such uses will be "relinquished" are highly unrealistic. Already, DOD is developing "smart dust," which are tiny robots the size of insects or even smaller. Although not quite nanotechnology, millions of these devices can be dropped into enemy territory to provide highly detailed surveillance. The potential application for even smaller, nanotechnology-based devices is even greater. Want to find Saddam Hussein or Osama bin Laden? Need to locate hidden weapons of mass destruction? Billions of essentially invisible spies could monitor every square inch of enemy territory, identify every person and every weapon, and even carry out missions to destroy enemy targets. The only way for an enemy to counteract such a force is, of course, with their own nanotechnology. The point is that nanotechnology-based weapons will obsolete weapons of larger size.

In addition, nanobots will also be able to expand our experiences and our capabilities. Nanobot technology will provide fully immersive, totally convincing virtual reality in the following way. The nanobots take up positions in close physical proximity to every interneuronal connection coming from all of our senses (e.g., eyes, ears, skin). We already have the technology for electronic devices to communicate with neurons in both directions that requires no direct physical contact with the neurons. For example, scientists at the Max Planck Institute have developed "neuron transistors" that can detect the firing of a nearby neuron, or alternatively, can cause a nearby neuron to fire, or suppress it from firing. This amounts to two-way communication between neurons and the electronic-based neuron transistors. The Institute scientists demonstrated their invention by controlling the movement of a living leech from their computer. Again, the primary aspect of nanobot-based virtual reality that is not yet feasible is size and cost.

When we want to experience real reality, the nanobots just stay in position (in the capillaries) and do nothing. If we want to enter virtual reality, they suppress all of the inputs coming from the real senses, and replace them with the signals that would be appropriate for the virtual environment. You (i.e., your brain) could decide to cause your muscles and limbs to move as you normally would, but the nanobots again intercept these interneuronal signals, suppress your real limbs from moving, and instead cause your virtual limbs to move and provide the appropriate movement and reorientation in the virtual environment.

The Web will provide a panoply of virtual environments to explore. Some will be recreations of real places, others will be fanciful environments that have no "real" counterpart. Some indeed would be impossible in the physical world (perhaps, because they violate the laws of physics). We will be able to "go" to these virtual environments by ourselves, or we will meet other people there, both real people and simulated people. Of course, ultimately there won't be a clear distinction between the two.

By 2030, going to a web site will mean entering a full-immersion virtual-reality environment. In addition to encompassing all of the senses, these shared environments can include emotional overlays as the nanobots will be capable of triggering the neurological correlates of emotions, sexual pleasure, and other derivatives of our sensory experience and mental reactions.

In the same way that people today beam their lives from web cams in their bedrooms, "experience beamers" circa 2030 will beam their entire flow of sensory experiences, and if so desired, their emotions and other secondary reactions. We'll be able to plug in (by going to the appropriate web site) and experience other people's lives as in the plot concept of 'Being John Malkovich.' Particularly interesting experiences can be archived and relived at any time.

We won't need to wait until 2030 to experience shared virtual-reality environments, at least for the visual and auditory senses. Full-immersion visual-auditory environments will be available by the end of this decade, with images written directly onto our retinas by our eyeglasses and contact lenses. All of the electronics for the computation, image reconstruction, and very high bandwidth wireless connection to the Internet will be embedded in our glasses and woven into our clothing, so computers as distinct objects will disappear.

In my view, the most significant implication of the development of nanotechnology and related advanced technologies of the 21st century will be the merger of biological and nonbiological intelligence. First, it is important to point out that well before the end of the twenty-first century, thinking on nonbiological substrates will dominate. Biological thinking is stuck at 10^{26} calculations per second (for all biological human brains), and that figure will not appreciably change, even with bioengineering changes to our genome. Nonbiological intelligence, on the other hand, is growing at a double-exponential rate and will vastly exceed biological intelligence well before the middle of this century. However, in my view, this nonbiological intelligence should still be considered human as it is fully derivative of the human-machine civilization. The merger of these two worlds of intelligence is not merely a merger of biological and nonbiological thinking mediums, but more importantly one of method and organization of thinking.

One of the key ways in which the two worlds can interact will be through nanobots. Nanobot technology will be able to expand our minds in virtually any imaginable way. Our brains today are relatively fixed in design. Although we do add patterns of interneuronal connections and neurotransmitter concentrations as a normal part of the learning process, the current overall capacity of the human brain is highly constrained, restricted to a mere hundred trillion connections. Brain implants based on massively distributed intelligent nanobots will ultimately expand our memories a trillion fold, and otherwise vastly improve all of our sensory, pattern recognition, and cognitive abilities. Since the nanobots are communicating with each other over a wireless local area network, they can create any set of new neural connections, can break existing connections (by suppressing neural firing), can create new hybrid biological-nonbiological networks, as well as add vast new nonbiological networks.

Using nanobots as brain extenders is a significant improvement over the idea of surgically installed neural implants, which are beginning to be used today (e.g., ventral posterior nucleus, subthalmic nucleus, and ventral lateral thalamus neural implants to counteract Parkinson's Disease and tremors from other neurological disorders, cochlear implants, and others.) Nanobots will be introduced without surgery, essentially just by injecting or even swallowing them. They can all be directed to leave, so the process is easily reversible. They are programmable, in that they can provide virtual reality one minute, and a variety of brain extensions the next. They can change their configuration, and clearly can alter their software. Perhaps most importantly, they are massively distributed and therefore can take up billions or trillions of positions throughout the brain, whereas a surgically introduced neural implant can only be placed in one or at most a few locations.

The Economic Imperatives of the Law of Accelerating Returns

It is the economic imperative of a competitive marketplace that is driving technology forward and fueling the law of accelerating returns. In turn, the law of accelerating returns is transforming economic relationships.

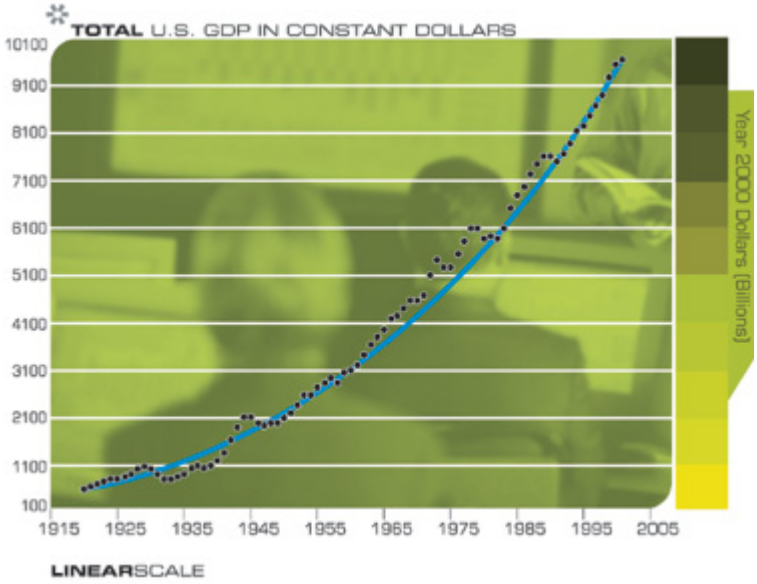
The primary force driving technology is economic imperative. We are moving towards nanoscale machines, as well as more intelligent machines, as the result of a myriad of small advances, each with their own particular economic justification.

To use one small example of many from my own experience at one of my companies (Kurzweil Applied Intelligence), whenever we came up with a slightly more intelligent version of speech recognition, the new version invariably had greater value than the earlier generation and, as a result, sales increased. It is interesting to note that in the example of speech recognition software, the three primary surviving competitors stayed very close to each other in the intelligence of their software. A few other companies that failed to do so (e.g., Speech Systems) went out of business. At any point in time, we would be able to sell the version prior to the latest version for perhaps a quarter of the price of the current version. As for versions of our technology that were two generations old, we couldn't even give those away.

There is a vital economic imperative to create smaller and more intelligent technology. Machines that can more precisely carry out their missions have enormous value. That is why they are being

built. There are tens of thousands of projects that are advancing the various aspects of the law of accelerating returns in diverse incremental ways. Regardless of near-term business cycles, the support for "high tech" in the business community, and in particular for software advancement, has grown enormously. When I started my optical character recognition (OCR) and speech synthesis company (Kurzweil Computer Products, Inc.) in 1974, high-tech venture deals totaled approximately $10 million. Even during today's high tech recession, the figure is 100 times greater. We would have to repeal capitalism and every visage of economic competition to stop this progression.

The economy (viewed either in total or per capita) has been growing exponentially throughout this century:

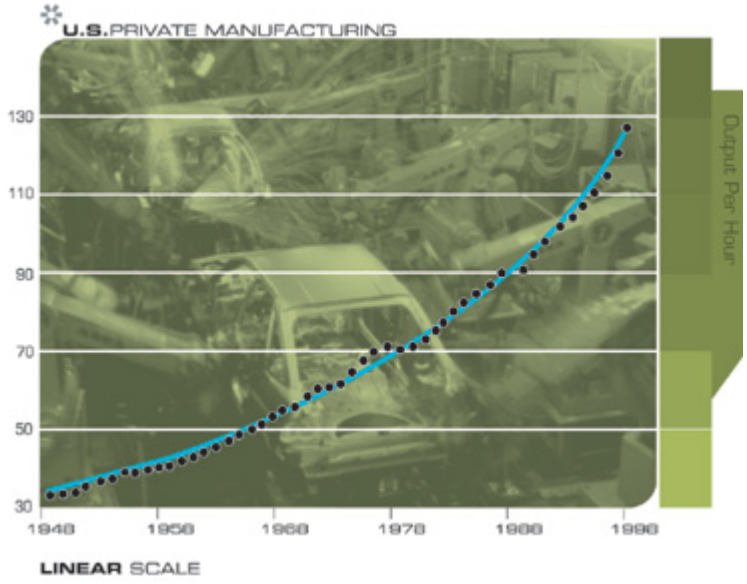


Note that the underlying exponential growth in the economy is a far more powerful force than periodic recessions. Even the "Great Depression" represents only a minor blip compared to the underlying pattern of growth. Most importantly, recessions, including the depression, represent only temporary deviations from the underlying curve. In each case, the economy ends up exactly where it would have been had the recession/depression never occurred.

Productivity (economic output per worker) has also been growing exponentially. Even these statistics are greatly understated because they do not fully reflect significant improvements in the quality and features of products and services. It is not the case that "a car is a car;" there have been significant improvements in safety, reliability, and features. Certainly, $1000 of computation today is immeasurably more powerful than $1000 of computation ten years ago (by a factor of more than1000). There are a myriad of such examples. Pharmaceutical drugs are increasingly effective. Products ordered in five minutes on the web and delivered to your door are worth more than products that you have to fetch yourself. Clothes custom-manufactured for your unique body scan are worth more than clothes you happen to find left on a store rack. These sorts of improvements are true for most product categories, and none of them are reflected in the productivity statistics.

The statistical methods underlying the productivity measurements tend to factor out gains by essentially concluding that we still only get one dollar of products and services for a dollar despite the fact that we get much more for a dollar (e.g., compare a $1,000 computer today to one ten years ago). University of Chicago Professor Pete Klenow and University of Rochester Professor Mark Bils estimate that the value of existing goods has been increasing at 1.5% per year for the past 20 years because of qualitative improvements. This still does not account for the introduction of entirely new products and product categories (e.g., cell phones, pagers, pocket computers). The Bureau of Labor Statistics, which is responsible for the inflation statistics, uses a model that incorporates an estimate of quality growth at only 0.5% per year, reflecting a systematic underestimate of quality improvement and a resulting overestimate of inflation by at least 1 percent per year.

Despite these weaknesses in the productivity statistical methods, the gains in productivity are now reaching the steep part of the exponential curve. Labor productivity grew at 1.6% per year until 1994, then rose at 2.4% per year, and is now growing even more rapidly. In the quarter ending July 30, 2000, labor productivity grew at 5.3%. Manufacturing productivity grew at 4.4% annually from 1995 to 1999, durables manufacturing at 6.5% per year.



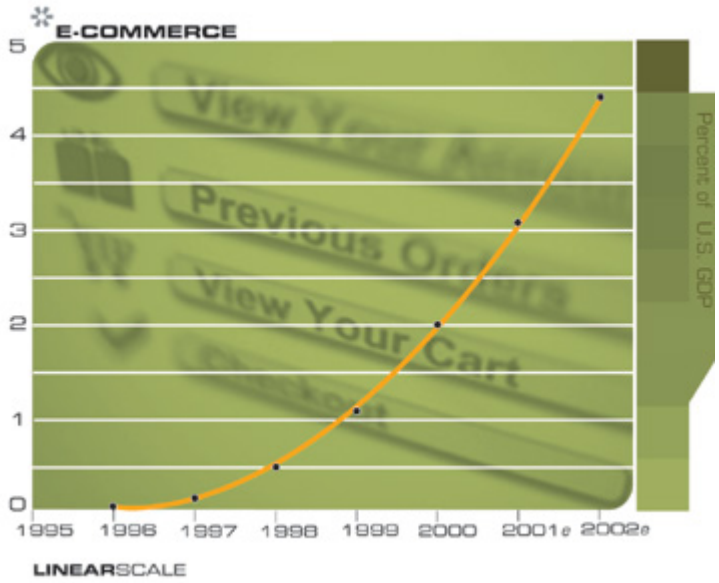
The 1990s have seen the most powerful deflationary forces in history. This is why we are not seeing inflation. Yes, it's true that low unemployment, high asset values, economic growth, and other such factors are inflationary, but these factors are offset by the double-exponential trends in the price-performance of all information-based technologies: computation, memory, communications, biotechnology, miniaturization, and even the overall rate of technical progress. These technologies deeply affect all industries. We are also undergoing massive disintermediation in the channels of distribution through the Web and other new communication technologies, as well as escalating efficiencies in operations and administration.

All of the technology trend charts above represent massive deflation. There are many examples of the impact of these escalating efficiencies. BP Amoco's cost for finding oil is now less than $1 per barrel, down from nearly $10 in 1991. Processing an Internet transaction costs a bank one penny, compared to over $1 using a teller ten years ago. A Roland Berger/Deutsche Bank study estimates a cost savings of $1200 per North American car over the next five years. A more optimistic Morgan Stanley study estimates that Internet-based procurement will save Ford, GM, and DaimlerChrysler about $2700 per vehicle.

It is important to point out that a key implication of nanotechnology is that it will bring the economics of software to hardware, i.e., to physical products. Software prices are deflating even more quickly than hardware.

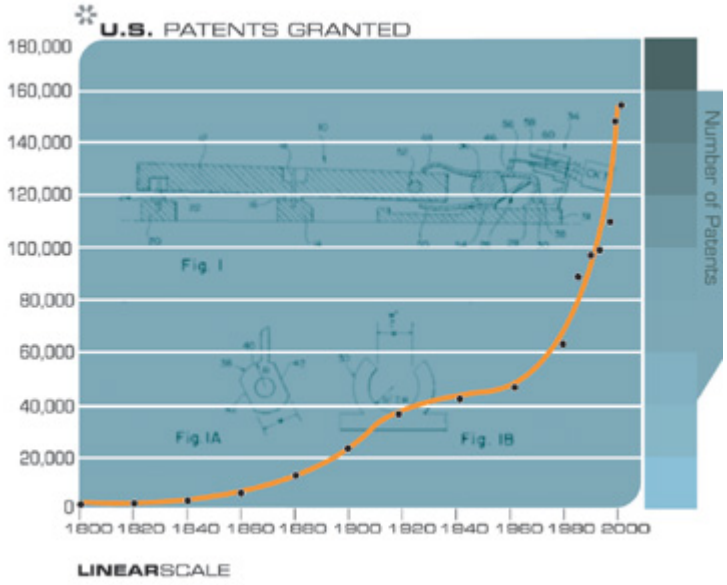
Software Price-Performance Has Also Improved at an Exponential Rate (Example: Automatic Speech Recognition Software)

	1985	**1995**	**2000**
Price	$5,000	$500	$50
Vocabulary Size (# words)	1,000	10,000	100,000
Continuous Speech?	No	No	Yes
User Training Required (Minutes)	180	60	5
Accuracy	Poor	Fair	Good



Current economic policy is based on outdated models that include energy prices, commodity prices, and capital investment in plant and equipment as key driving factors, but do not adequately model the size of technology, bandwidth, MIPs, megabytes, intellectual property, knowledge, and other increasingly vital (and increasingly increasing) constituents that are driving the economy.

Another indication of the law of accelerating returns in the exponential growth of human knowledge, including intellectual property. If we look at the development of intellectual property within the nanotechnology field, we see even more rapid growth.



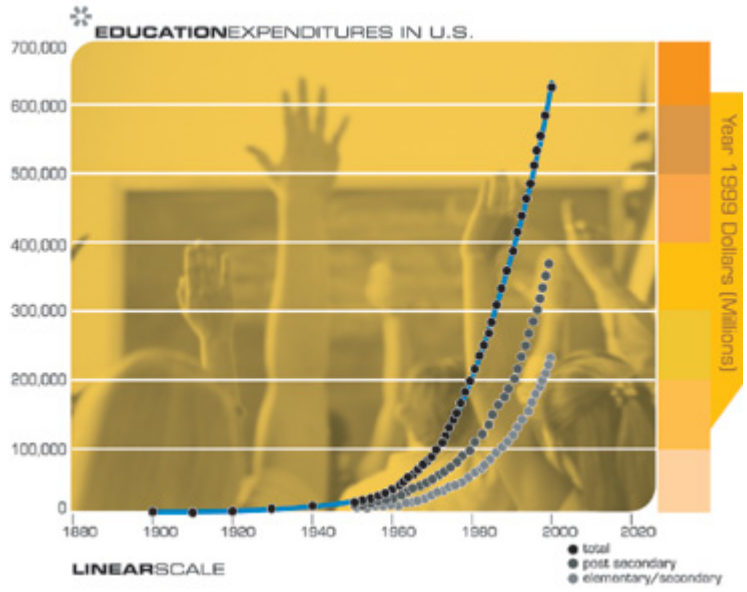
None of this means that cycles of recession will disappear immediately. Indeed there is a current economic slowdown and a technology-sector recession. The economy still has some of the underlying dynamics that historically have caused cycles of recession, specifically excessive commitments such as over-investment, excessive capital intensive projects and the overstocking of inventories. However, the rapid dissemination of information, sophisticated forms of online procurement, and increasingly transparent markets in all industries have diminished the impact of this cycle. So "recessions" are likely to have less direct impact on our standard of living. The underlying long-term growth rate will continue at a double exponential rate.

Moreover, innovation and the rate of paradigm shift are not noticeably affected by the minor deviations caused by economic cycles. All of the technologies exhibiting exponential growth shown in the above charts are continuing without losing a beat through this economic slowdown.

The overall growth of the economy reflects completely new forms and layers of wealth and value that did not previously exist, or least that did not previously constitute a significant portion of the economy (but do now): new forms of nanoparticle-based materials, genetic information, intellectual property, communication portals, web sites, bandwidth, software, data bases, and many other new technology-based categories.

Another implication of the law of accelerating returns is exponential growth in education and learning. Over the past 120 years, we have increased our investment in K-12 education (per student and in constant dollars) by a factor of ten. We have a one hundred fold increase in the number of college students. Automation started by amplifying the power of our muscles, and in recent times has been amplifying the power of our minds. Thus, for the past two centuries, automation has been eliminating jobs at the bottom of the skill ladder while creating new (and

better paying) jobs at the top of the skill ladder. So the ladder has been moving up, and thus we have been exponentially increasing investments in education at all levels.



The Deeply Intertwined Promise and Peril of Nanotechnology and Related Advanced Technologies

Technology has always been a double-edged sword, bringing us longer and healthier life spans, freedom from physical and mental drudgery, and many new creative possibilities on the one hand, while introducing new and salient dangers on the other. Technology empowers both our creative and destructive natures. Stalin's tanks and Hitler's trains used technology. We still live today with sufficient nuclear weapons (not all of which appear to be well accounted for) to end all mammalian life on the planet. Bioengineering is in the early stages of enormous strides in reversing disease and aging processes. However, the means and knowledge will soon exist in a routine college bioengineering lab (and already exists in more sophisticated labs) to create unfriendly pathogens more dangerous than nuclear weapons. As technology accelerates towards the full realization of biotechnology, nanotechnology and "strong" AI (artificial intelligence at human levels and beyond), we will see the same intertwined potentials: a feast of creativity resulting from human intelligence expanded many-fold combined with many grave new dangers.

Consider unrestrained nanobot replication. Nanobot technology requires billions or trillions of such intelligent devices to be useful. The most cost-effective way to scale up to such levels is through self-replication, essentially the same approach used in the biological world. And in the same way that biological self-replication gone awry (i.e., cancer) results in biological destruction, a defect in the mechanism curtailing nanobot self-replication would endanger all physical entities, biological or otherwise. I address below steps we can take to address this grave risk, but we cannot have complete assurance in any strategy that we devise today.

Other primary concerns include "who is controlling the nanobots?" and "who are the nanobots talking to?" Organizations (e.g., governments, extremist groups) or just a clever individual could put trillions of undetectable nanobots in the water or food supply of an individual or of an entire population. These "spy" nanobots could then monitor, influence, and even control our thoughts and actions. In addition to introducing physical spy nanobots, existing nanobots could be influenced through software viruses and other software "hacking" techniques. When there is software running in our brains, issues of privacy and security will take on a new urgency.

My own expectation is that the creative and constructive applications of this technology will dominate, as I believe they do today. However, I believe we need to invest more heavily in developing specific defensive technologies. As I address further below, we are at this stage today for biotechnology, and will reach the stage where we need to directly implement defensive technologies for nanotechnology during the late teen years of this century.

If we imagine describing the dangers that exist today to people who lived a couple of hundred years ago, they would think it mad to take such risks. On the other hand, how many people in the year 2000 would really want to go back to the short, brutish, disease-filled, poverty-stricken, disaster-prone lives that 99 percent of the human race struggled through a couple of centuries ago? We may romanticize the past, but up until fairly recently, most of humanity lived extremely fragile lives where one all-too-common misfortune could spell disaster. Substantial portions of our species still live in this precarious way, which is at least one reason to continue technological progress and the economic enhancement that accompanies it.

People often go through three stages in examining the impact of future technology: awe and wonderment at its potential to overcome age old problems; then a sense of dread at a new set of grave dangers that accompany these new technologies; followed, finally and hopefully, by the realization that the only viable and responsible path is to set a careful course that can realize the promise while managing the peril.

This congressional hearing was party inspired by Bill Joy's cover story for Wired magazine, *Why The Future Doesn't Need* Us. Bill Joy, cofounder of Sun Microsystems and principal developer of the Java programming language, has recently taken up a personal mission to warn us of the impending dangers from the emergence of self-replicating technologies in the fields of genetics, nanotechnology, and robotics, which he aggregates under the label "GNR." Although his warnings are not entirely new, they have attracted considerable attention because of Joy's credibility as one of our leading technologists. It is reminiscent of the attention that George Soros, the currency arbitrager and arch capitalist, received when he made vaguely critical comments about the excesses of unrestrained capitalism.

Joy's concerns include genetically altered designer pathogens, followed by self-replicating entities created through nanotechnology. And if we manage to survive these first two perils, we will encounter robots whose intelligence will rival and ultimately exceed our own. Such robots may make great assistants, but who's to say that we can count on them to remain reliably friendly to mere humans?

Although I am often cast as the technology optimist who counters Joy's pessimism, I do share his concerns regarding self-replicating technologies; indeed, I played a role in bringing these dangers to Bill's attention. In many of the dialogues and forums in which I have participated on this subject, I end up defending Joy's position with regard to the feasibility of these technologies and scenarios when they come under attack by commentators who I believe are being quite shortsighted in their skepticism. Even so, I do find fault with Joy's prescription: halting the advance of technology and the pursuit of knowledge in broad fields such as nanotechnology.

In his essay, Bill Joy eloquently described the plagues of centuries past and how new self-replicating technologies, such as mutant bioengineered pathogens and "nanobots" run amok, may bring back long-forgotten pestilence. Indeed these are real dangers. It is also the case, which Joy acknowledges, that it has been technological advances, such as antibiotics and improved sanitation, which have freed us from the prevalence of such plagues. Suffering in the world continues and demands our steadfast attention. Should we tell the millions of people afflicted with cancer and other devastating conditions that we are canceling the development of all bioengineered treatments because there is a risk that these same technologies may someday be used for malevolent purposes? Having asked the rhetorical question, I realize that there is a movement to do exactly that, but I think most people would agree that such broad-based relinquishment is not the answer.

The continued opportunity to alleviate human distress is one important motivation for continuing technological advancement. Also compelling are the already apparent economic gains I discussed above that will continue to hasten in the decades ahead. The continued acceleration of many intertwined technologies are roads paved with gold (I use the plural here because technology is clearly not a single path). In a competitive environment, it is an economic imperative to go down these roads. Relinquishing technological advancement would be economic suicide for individuals, companies, and nations.

The Relinquishment Issue

This brings us to the issue of relinquishment, which is Bill Joy's most controversial recommendation and personal commitment. I do feel that relinquishment at the right level is part of a responsible and constructive response to these genuine perils. The issue, however, is exactly this: at what level are we to relinquish technology?

Ted Kaczynski would have us renounce all of it. This, in my view, is neither desirable nor feasible, and the futility of such a position is only underscored by the senselessness of Kaczynski's deplorable tactics. There are other voices, less reckless than Kaczynski, who are nonetheless arguing for broad-based relinquishment of technology. Bill McKibben, the environmentalist who was one of the first to warn against global warming, takes the position that "environmentalists must now grapple squarely with the idea of a world that has enough wealth and enough technological capability, and should not pursue more." In my view, this position ignores the extensive suffering that remains in the human world, which we will be in a position to alleviate through continued technological progress.

Another level would be to forego certain fields — nanotechnology, for example — that might be regarded as too dangerous. But such sweeping strokes of relinquishment are equally untenable. As I pointed out above, nanotechnology is simply the inevitable end result of the persistent trend towards miniaturization that pervades all of technology. It is far from a single centralized effort, but is being pursued by a myriad of projects with many diverse goals.

One observer wrote:

"A further reason why industrial society cannot be reformed . . . is that modern technology is a unified system in which all parts are dependent on one another. You can't get rid of the "bad" parts of technology and retain only the "good" parts. Take modern medicine, for example. Progress in medical science depends on progress in chemistry, physics, biology, computer science and other fields. Advanced medical treatments require expensive, high-tech equipment that can be made available only by a technologically progressive, economically rich society. Clearly you can't have much progress in medicine without the whole technological system and everything that goes with it."

The observer I am quoting is, again, Ted Kaczynski. Although one will properly resist Kaczynski as an authority, I believe he is correct on the deeply entangled nature of the benefits and risks. However, Kaczynski and I clearly part company on our overall assessment on the relative balance between the two. Bill Joy and I have dialogued on this issue both publicly and privately, and we both believe that technology will and should progress, and that we need to be actively concerned with the dark side. If Bill and I disagree, it's on the granularity of relinquishment that is both feasible and desirable.

Abandonment of broad areas of technology will only push them underground where development would continue unimpeded by ethics and regulation. In such a situation, it would be the less-stable, less-responsible practitioners (e.g., terrorists) who would have all the expertise.

I do think that relinquishment at the right level needs to be part of our ethical response to the dangers of 21st century technologies. One constructive example of this is the proposed ethical guideline by the Foresight Institute, founded by nanotechnology pioneer Eric Drexler, that nanotechnologists agree to relinquish the development of physical entities that can self-replicate in a natural environment. Another is a ban on self-replicating physical entities that contain their own codes for self-replication. In what nanotechnologist Ralph Merkle calls the "broadcast architecture," such entities would have to obtain such codes from a centralized secure server, which would guard against undesirable replication. I discuss these guidelines further below.

The broadcast architecture is impossible in the biological world, which represents at least one way in which nanotechnology can be made safer than biotechnology. In other ways, nanotech is potentially more dangerous because nanobots can be physically stronger than protein-based entities and more intelligent. It will eventually be possible to combine the two by having nanotechnology provide the codes within biological entities (replacing DNA), in which case biological entities can use the much safer broadcast architecture. I comment further on the strengths and weaknesses of the broadcast architecture below.

As responsible technologies, our ethics should include such "fine-grained" relinquishment, among other professional ethical guidelines. Other protections will need to include oversight by regulatory bodies, the development of technology-specific "immune" responses, as well as computer assisted surveillance by law enforcement organizations. Many people are not aware that our intelligence agencies already use advanced technologies such as automated word spotting to monitor a substantial flow of telephone conversations. As we go forward, balancing our cherished rights of privacy with our need to be protected from the malicious use of powerful 21st century technologies will be one of many profound challenges. This is one reason that such issues as an encryption "trap door" (in which law enforcement authorities would have access to otherwise secure information) and the FBI "Carnivore" email-snooping system have been controversial, although these controversies have abated since 9-11-2001.

As a test case, we can take a small measure of comfort from how we have dealt with one recent technological challenge. There exists today a new form of fully nonbiological self replicating entity that didn't exist just a few decades ago: the computer virus. When this form of destructive intruder first appeared, strong concerns were voiced that as they became more sophisticated, software pathogens had the potential to destroy the computer network medium they live in. Yet the "immune system" that has evolved in response to this challenge has been largely effective. Although destructive self-replicating software entities do cause damage from time to time, the injury is but a small fraction of the benefit we receive from the computers and communication links that harbor them. No one would suggest we do away with computers, local area networks, and the Internet because of software viruses.

One might counter that computer viruses do not have the lethal potential of biological viruses or of destructive nanotechnology. This is not always the case; we rely on software to monitor patients in critical care units, to fly and land airplanes, to guide intelligent weapons in our current campaign in Iraq, and other "mission-critical" tasks. To the extent that this is true, however, this observation only strengthens my argument. The fact that computer viruses are not usually deadly to humans only means that more people are willing to create and release them. It also means that our response to the danger is that much less intense. Conversely, when it comes to self-replicating entities that are potentially lethal on a large scale, our response on all levels will be vastly more serious, as we have seen since 9-11.

I would describe our response to software pathogens as effective and successful. Although they remain (and always will remain) a concern, the danger remains at a nuisance level. Keep in mind that this success is in an industry in which there is no regulation, and no certification for practitioners. This largely unregulated industry is also enormously productive. One could argue that it has contributed more to our technological and economic progress than any other enterprise in human history. I discuss the issue of regulation further below.

Development of Defensive Technologies and the Impact of Regulation

Joy's treatise is effective because he paints a picture of future dangers as if they were released on today's unprepared world. The reality is that the sophistication and power of our defensive technologies and knowledge will grow along with the dangers. When we have "gray goo" (unrestrained nanobot replication), we will also have "blue goo" ("police" nanobots that combat

the "bad" nanobots). The story of the 21st century has not yet been written, so we cannot say with assurance that we will successfully avoid all misuse. But the surest way to prevent the development of the defensive technologies would be to relinquish the pursuit of knowledge in broad areas. We have been able to largely control harmful software virus replication because the requisite knowledge is widely available to responsible practitioners. Attempts to restrict this knowledge would have created a far less stable situation. Responses to new challenges would have been far slower, and it is likely that the balance would have shifted towards the more destructive applications (e.g., software viruses).

The challenge most immediately in front of us is not self-replicating nanotechnology, but rather self-replicating biotechnology. The next two decades will be the golden age of biotechnology, whereas the comparable era for nanotechnology will follow in the 2020s and beyond. We are now in the early stages of a transforming technology based on the intersection of biology and information science. We are learning the "software" methods of life and disease processes. By reprogramming the information processes that lead to and encourage disease and aging, we will have the ability to overcome these afflictions. However, the same knowledge can also empower a terrorist to create a bioengineered pathogen.

As we compare the success we have had in controlling engineered software viruses to the coming challenge of controlling engineered biological viruses, we are struck with one salient difference. As I noted above, the software industry is almost completely unregulated. The same is obviously not the case for biotechnology. A bioterrorist does not need to put his "innovations" through the FDA. However, we do require the scientists developing the defensive technologies to follow the existing regulations, which slow down the innovation process at every step. Moreover, it is impossible, under existing regulations and ethical standards, to test defenses to bioterrorist agents. There is already extensive discussion to modify these regulations to allow for animal models and simulations to replace infeasible human trials. This will be necessary, but I believe we will need to go beyond these steps to accelerate the development of vitally needed defensive technologies.

For reasons I have articulated above, stopping these technologies is not feasible, and pursuit of such broad forms of relinquishment will only distract us from the vital task in front of us. In terms of public policy, the task at hand is to rapidly develop the defensive steps needed, which include ethical standards, legal standards, and defensive technologies. It is quite clearly a race. As I noted, in the software field, the defensive technologies have remained a step ahead of the offensive ones. With the extensive regulation in the medical field slowing down innovation at each stage, we cannot have the same confidence with regard to the abuse of biotechnology.

In the current environment, when one person dies in gene therapy trials, there are congressional investigations and all gene therapy research comes to a temporary halt. There is a legitimate need to make biomedical research as safe as possible, but our balancing of risks is completely off. The millions of people who desperately need the advances to be made available by gene therapy and other breakthrough biotechnology advances appear to carry little political weight against a handful of well-publicized casualties from the inevitable risks of progress.

This equation will become even more stark when we consider the emerging dangers of bioengineered pathogens. What is needed is a change in public attitude in terms of tolerance for needed risk.

Hastening defensive technologies is absolutely vital to our security. We need to streamline regulatory procedures to achieve this. However, we also need to greatly increase our investment explicitly in the defensive technologies. In the biotechnology field, this means the rapid development of antiviral medications. We will not have time to develop specific countermeasures for each new challenge that comes along. We are close to developing more generalized antiviral technologies, and these need to be accelerated.

I have addressed here the issue of biotechnology because that is the threshold and challenge that we now face. The comparable situation will exist for nanotechnology once replication of nano-engineered entities has been achieved. As that threshold comes closer, we will then need to invest specifically in the development of defensive technologies, including the creation of a nanotechnology-based immune system. Bill Joy and other observers have pointed out that such an immune system would itself be a danger because of the potential of "autoimmune" reactions (i.e., the immune system using its powers to attack the world it is supposed to be defending).

However, this observation is not a compelling reason to avoid the creation of an immune system. No one would argue that humans would be better off without an immune system because of the possibility of auto immune diseases. Although the immune system can itself be a danger, humans would not last more than a few weeks (barring extraordinary efforts at isolation) without one. The development of a technological immune system for nanotechnology will happen even without explicit efforts to create one. We have effectively done this with regard to software viruses. We created a software virus immune system not through a formal grand design project, but rather through our incremental responses to each new challenge. We can expect the same thing will happen as challenges from nanotechnology based dangers emerge. The point for public policy will be to specifically invest in these defensive technologies.

It is premature today to develop specific defensive nanotechnologies since we can only have a general idea of what we are trying to defend against. It would be similar to the engineering world creating defenses against software viruses before the first one had been created. However, there is already fruitful dialogue and discussion on anticipating this issue, and significantly expanded investment in these efforts is to be encouraged.

As I mentioned above, the Foresight Institute, for example, has devised a set of ethical standards and strategies for assuring the development of safe nanotechnology. These guidelines include:

- "Artificial replicators must not be capable of replication in a natural, uncontrolled environment."
- "Evolution within the context of a self-replicating manufacturing system is discouraged."
- "MNT (molecular nanotechnology) designs should specifically limit proliferation and provide traceability of any replicating systems."

- "Distribution of molecular manufacturing development capability should be restricted whenever possible, to responsible actors that have agreed to the guidelines. No such restriction need apply to end products of the development process."

Other strategies that the Foresight Institute has proposed include:

- Replication should require materials not found in the natural environment.
- Manufacturing (replication) should be separated from the functionality of end products. Manufacturing devices can create end products, but cannot replicate themselves, and end products should have no replication capabilities.
- Replication should require replication codes that are encrypted, and time limited. The broadcast architecture mentioned earlier is an example of this recommendation.

These guidelines and strategies are likely to be effective with regarding to preventing accidental release of dangerous self-replicating nanotechnology entities. The situation with regard to intentional design and release of such entities is more complex and more challenging. We can anticipate approaches that would have the potential to defeat each of these layers of protections by a sufficiently determined and destructive opponent.

Take, for example, the broadcast architecture. When properly designed, each entity is unable to replicate without first obtaining replication codes. These codes are not passed on from one replication generation to the next. However, a modification to such a design could bypass the destruction of the replication codes and thereby pass them on to the next generation. To overcome that possibility, it has been recommended that the memory for the replication codes be limited to only a subset of the full replication code so that there is insufficient memory to pass the codes along. However, this guideline could be defeated by expanding the size of the replication code memory to incorporate the entire code. Another protection that has been suggested is to encrypt the codes and to build in protections such as time expiration limitations in the decryption systems. However, we can see the ease with which protections against unauthorized replications of intellectual property such as music files has been defeated. Once replication codes and protective layers are stripped away, the information can be replicated without these restrictions.

My point is not that protection is impossible. Rather, we need to realize that any level of protection will only work to a certain level of sophistication. The "meta" lesson here is that we will need to continue to advance the defensive technologies, and keep them one or more steps ahead of the destructive technologies. We have seen analogies to this in many areas, including technologies for national defense, as well as our largely successful efforts to combat software viruses, that I alluded to above.

What we can do today with regard to the critical challenge of self-replication in nanotechnology is to continue the type of effective study that the Foresight Institute has initiated. With the human genome project, three to five percent of the budgets were devoted to the ethical, legal, and social implications (ELSI) of the technology. A similar commitment for nanotechnology would be appropriate and constructive.

Technology will remain a double-edged sword, and the story of the 21st century has not yet been written. It represents vast power to be used for all humankind's purposes. We have no choice but to work hard to apply these quickening technologies to advance our human values, despite what often appears to be a lack of consensus on what those values should be.