THE RETURN OF THE KRELL MACHINE

Nanotechnology, the Singularity, and the Empty Planet Syndrome

by Steven B. Harris [1]
© 2001, 2002

What will happen when humans gain the ability to manufacture nearly anything we want, and when our machines surpass our own intelligence? We had better hope the results are better than we see in science fiction, because, in a few generations, both these situations may well be upon us.

I. Introduction: *Forbidden Planet and the Ultimate Machine*

In 1956, the Fred McLeod Wilcox film *Forbidden Planet* became the second memorable science fiction movie of the 1950's (the first being Robert Wise's *The Day The Earth Stood Still*). *Forbidden Planet*, from a screenplay by Cyril Hume, is still entertaining today. It has become a classic by being among the first films to raise important issues about the use of ultimate technologies. Moreover, it has also had a vast impact on the art of science fiction films which followed it.

Modern viewers of *Forbidden Planet* are reminded of *Star Trek*, but of course the connection is in the other direction. Many episodes of *Trek* borrow liberally from *Forbidden Planet*. As the film begins, a "United Planets Cruiser," featuring a dashing young starship captain, is paying a call to the planet Altair IV, to investigate the loss of a science mission there 20 years before. They find no one alive on the planet save for the expedition's strangely powerful philologist, one Edward Morbius, Ph.D. (lit.), and his intriguing and beautiful teenaged daughter, who has never seen humans other than her father. (We recognize the basic plot of *The Tempest* from Shakespeare, of *Star Trek's* episode *Requiem for Methuselah*, and many others. The captain is in for trouble.) Dr. Morbius, attended by an advanced robot servant, is engaged in solo decipherment of traces of an alien civilization which had once occupied the planet, but which had become suddenly extinct 200,000 years before. In a key scene, Morbius, in almost blank verse, tells the starship captain about this vanished race, which had called themselves the ***Krell*:**

Ethically, as well as technologically,
  they were a million years ahead of humankind.
For, in unlocking the mysteries of nature,
  they had conquered even their baser-selves.

And, when in the course of eons,
  they had abolished sickness and insanity

and crime and all injustice,
they turned, still with high benevolence,
outward toward space.

Long before the dawn of man's history,
they had walked our Earth,
and brought back many biological specimens.

The heights they had reached!

But then--- seemingly on the threshold
of some supreme accomplishment
which was to have crowned their entire history--
this all-but-divine race perished,
in a single night.

In the two thousand centuries
since that unexplained catastrophe,
even their cloud-piercing towers
of glass and porcelain and adamantine steel
have crumbled back into the soil of Altair IV,
and nothing, absolutely nothing,
remains aboveground.

Later, Morbius shows the starship captain the principal remains of the Krell civilization: a self-repairing and still-functioning gigantic machine which reposes, blinking and humming, beneath an empty desert of Altair IV. It is a cube measuring 20 miles on a side (think "Borg Cube" from *Star Trek*) powered by 9,200 working thermonuclear (fusion) reactors. Its function is a mystery, but later is finally revealed. The huge device was built by the Krell as a replacement for all technological instrumentalities. It is a technical Aladdin's lamp, an Ultimate Machine waiting for a command. The starship captain finally figures this out, with some clues from the brain-boosted (and brain-burned) ship's doctor, and accosts Dr. Morbius with the answer:

“Morbius— a big machine, 8000 cubic miles of klystron relays, enough power for a whole population of creative geniuses—operated by remote control! Morbius--- operated by the electromagnetic impulses of individual Krell brains. [..] In return, that machine would instantaneously project solid matter to any point on the planet. In any shape or color they might imagine. For any purpose, Morbius! Creation by pure thought!”

But there's also a little problem with such a technology, the captain tells Morbius; it is *Monsters From the Id:*

"But, like you, the Krell forgot one deadly danger-- their own subconscious hate and lust for destruction! [..] And so, those mindless beasts of the subconscious had access to a machine that could NEVER be shut down! The secret devil of every soul on the planet, all set free at once, to loot and maim! And take revenge, Morbius, and kill!"

The nightmare monsters from the machine allow the Krell to destroy themselves, and later (guided unwillingly now by Morbius' subconscious) the device acts as facilitator to destroy one human expedition and part of another. In the end, a desperate Morbius puts the machine into overload as a final stop to the invincible monsters (we see this scene later in the film *Alien*). The starship captain and Morbius' daughter manage to get away from Altair IV just in time before the planet explodes. Wiping out everything is what these ultimate machines all seem to do **[2].**

From our 21st century vantage-point, we recognize the Krell Machine as perhaps a 1950's metaphor for the relatively new nuclear energy— a technology thought at that time to be potentially a nearly infinite power source, for either good or evil. The question asked in the film is thus the famous one of this early atomic era: Are our Freudian Ids, our ape's-emotional-brains, *ready* for that kind of increase in power? If a machine had the power to instantly make for us anything we wanted, would we be wise enough to know what was good for us? The answer of *Forbidden Planet* is no.

But it's a temptation. Since *Forbidden Planet*, the Krell Machine has turned up repeatedly in science fiction, from *Star Trek* to *Total Recall*. Perhaps the most interesting set of ideas it prefigures is a group of now serious predictions about our future. It turns out that the Bomb is only a small subset of mankind's worst coming worries. A nuclear bomb, after all, is merely one more device we made when we grew smart enough to do it. The underlying problem is that we're getting both smarter and better at making things. Further, both of these trends are snowballing toward an inevitable avalanche, due to the fact that *each is starting to amplify the other.*

II. Mankind's Pending Ultimate Instrumentalities, Part A: Nanotechnology

Let us look now at the darkest potentials of foreseeable technology. The rule we set for ourselves is that we will not consider "fantasy" ideas, such as what may be possible if we discover new loopholes in physical laws. We wish merely to ask how far ordinary human technology may go, given known physical constraints. Such possible *ultimate technologies*, as we have suggested above, divide broadly into those connected with the physical world, and those connected with the mental and computational world.

We begin with the physical. Here, we are amused by one of the more advanced capabilities of Robby the Robot, who is the servant of Morbius in the 1956 film. Robby (a techno-version of

The Tempest's slave-spirit Ariel) is human-designed, using bits of advanced Krell knowledge. Robby can synthesize artificial gems of large size, and can analyze and duplicate any food or chemical mixture, all within the small space of his body. At one point we see Robby obligingly make 50 gallons of bootleg liquor for the starship's cook, who plays *The Tempest*'s drunken crewman/fool. Does any technology which we might realistically imagine, allow such powers?

We do not know the inspiration of physicist Richard Feynman, when he gave the answer to this question just three years later, in his now-famous essay *There's Plenty of Room at the Bottom* **[3]**. But perhaps part of the inspiration was this film. Feynman's answer in 1959 was surprising: the idea of total molecular-level materials manufacturing control may be science fiction, but it is far from fantasy. Feynman advised that there do not appear to be any physical laws which prohibit the manipulation and manufacture of things atom-by-atom, allowing (for example) the kinds of gem-synthesis and duplication of foodstuffs that Robby the Robot does.

In 1986, (***Engines of Creation***), K. Eric Drexler predicted some design details in a popular book. Complex chemical syntheses, he proposed, might be done using sub-microscopic construction-machines. Such machines (called “*assemblers*”) would work like natural biological catalysts (enzymes). By the time of Drexler's writing, it was known that enzymes work semi-mechanistically, using tiny chemically-powered protein "arms" to grab and move groups of atoms, changing the chemical bonds between them. (A chemical bond is a place where electrons are shared between atoms, causing the assembly to stick together to form a molecule). Drexler now proposed that assemblers, unlike most enzymes, would be *programmable*. Instead of only one chemical job, an assembler might be programmed to do many.

In Drexler's scheme, one could give a general-purpose assembler *instructions* about what types of atoms and bonds to look for and work on, changing these instructions as the device moved from one part of a molecule to the next. Fully programmable assemblers would thus have the full flexibility of computer-controlled industrial robots, but be able to use it on the size-scale of chemistry.

The potential power of such devices is already partly illustrated for us by the very fine synthetic detail seen in biology. In living systems a semi-programmable enzyme-complex called the *ribosome* is able to manufacture a potentially infinite number of different proteins (including enzymes for more ribosomes), using programming information on the fly from an “instruction tape” of messenger RNA. Drexler’s proposed devices, by analogy with the ribosome, would be more powerful and flexible still—able to take a much wider variety of instructions, and be able to make more complex decisions as they worked. Such devices would be able to make not only proteins, but any chemical structure that was stable.

Since Drexler’s proposal, some progress has been made. In 1989 scientists working for IBM used a very pointy needle to nudge 35 individual xenon atoms on a cold surface into spelling out "IBM" in letters a few atoms long. In 1996, further studies showed that molecules could be individually positioned, even at room temperature **[4]**. Thus, the crucial hurdle is not in manipulating individual atoms or molecules (this can be done) but in doing it cleverly enough.

We see immediately that there is a chicken-and-egg problem here. Cell-sized computers for running assemblers would be possible to construct *if* molecular-scale engineering capability was available to begin with. If not, the difficulty would lie in making the first assemblers. These would need to arise from a laborious process of miniaturizing manufacturing capability, level by level, to make the next smaller generation of devices, until we reached the molecule-sized bottom of chemical reality. Once devices were manufactured this small, however, things would become much easier. The assemblers would then be programmable to simply make more of themselves, just as living cells replicate their own ribosomes, and thus replicate themselves.

Nanotechnology (as Drexler referred to his program) would offer the ultimate physical manufacturing technology. Such manufacture would start with basic shapes. Josh Storrs Hall has proposed that nanomachines (“nanites”) of approximately protozoan size might interact tactilely with each other, to generate ordinary objects having low densities but high strengths. Solid objects might thus emerge from fluid dispersions like today's plastic *stereolithography* sculpture, yet at the same time potentially be as mobile and protean as the "liquid metal" automaton in the film *Terminator 2*. A collection of nanites might float like mist, but morph or solidify when instructed to lock arms. Such a “*Utility Fog*” would quickly become any shape or color we wish. Say the word, for example, and an extra chair might coalesce and shape itself out of mist which is otherwise nearly invisible. If you can do such deeds just by thinking or visualizing, you will be approaching Krell territory.

A notable application of nanotechnology would lie in its role as the ultimate medical treatment. Feynman reported in 1959 that his friend Al Hibbs had remarked, on hearing of tiny machines, that it would be very convenient to simply "swallow the doctor." Of course, the micro-doctor, working quickly and by touch, would need to have considerable on-site "intelligence." As early as 1949, science fiction author Hal Clement (***Needle***, first serialized in *Astounding SF*) had already sketched the regenerative possibilities if a human body were interpenetrated by an amorphous intelligent Being made of very tiny parts, which could sense and fix problems micro-surgically. Such beings are science fiction, but seem physically possible. The direct miniaturization of humans or their craft as seen in *Fantastic Voyage* is fantasy*,* for it requires the miniaturization of atoms, which is far outside the limits of known physical laws. But not so, the kinds of things that "inside doctoring" might do, if only the “doctor” were an intelligent but microscopic robot built of ordinary atoms, cleverly assembled. Normal atoms appear plenty small enough to make an intelligent machine far smaller than the human cells it may be tasked to repair [5].

Nanotechnology would not necessarily need to work inside a body to make biomaterials. It should be able to synthesize healthy tissue at any place, for any purpose. Proteins, cells, and tissues could be laid down in Utility Fog shaped forms. With the proper supply of information and raw materials, nanites might use an artificial circulatory system to manufacture and place cells on organ-shaped scaffolding. There would be no reason such an enterprise could not eventually manufacture a complete living organism.

With such biological manufacture, we come naturally to the most dramatic use of nanotechnology, which is the ability to duplicate and "fax" living organisms, including humans, using information taken (perhaps non-destructively) from a living template organism.

Living organisms as we know them now are constructed (we say "grown") slowly from the raw materials of simple food molecules, using a seed of information which controls some nanomachine-like cellular organelles (ribosomes, etc.). Nothing, however, stands in the way of improving this natural process greatly, in both rate and fidelity. The cellular clones of today are far from exact copies of the original organism, because DNA contains too little information for that. DNA is a *recipe*, not a blueprint. By contrast, nanotechnology in theory might read out the more complex "blueprint" of an existing individual human, use this far larger instruction set to build another *exact* copy. Something much more than a standard clone, which is only as interesting as an identical twin.

Moreover, rather than producing an adult human in 20 years, it might be possible to do it in months or weeks, including structure from a template brain so that memories and learning could be replicated also. Thus, while simple cellular cloning of humans *per se* will not be capable of presenting the kinds of social problems seen in the recent Schwarzenegger film *The Sixth Day* (2001), a fully *duplicative* nanotechnology *would* be up to the task. To be sure, a nanotechnologically-duplicated person might not quite pop into existence nearly so quickly as a matter-transportee on *Star Trek.* A human synthesis would also need machinery as well as raw materials in place at the "destination" point (the machinery could be grown on-site as well, from a small seed and instructions). These are details. The point is that the basic process, as well as all the ethical and philosophical problems attendant with it, does not seem to be ruled out by any physical laws we know **[6]**.

As we have hinted, however, the powers being discussed are not unlimited. Nanomachines are precision programmable chemical catalysts that are held together by chemical bonds, subject to standard inter- and intra-molecular forces. This places severe limits on the kinetic energy that machine pieces may have, and thus how fast they may work in order to move and assemble atoms. There is friction to deal with, molecular degradation, and of course the need for constant error correction, as in any complex system. There are also temperature and pressure constraints, again because nanomachines are made of ordinary molecular substances.

Further, nanotechnology techniques will have power over chemistry only; no nuclear transformations are included, so we cannot turn lead into gold. These are fundamental limitations connected with physical law, and not likely to be circumventable. Nanotechnology provides the limiting technology for how to make any chemically possible structure of atoms, on any scale that is stable. In theory, so long as feedstocks of chemical elements are available, one can duplicate any object that already exists in the relatively low-temperature and low-pressure part of our universe (i.e., at least crusts of small planets), though it won't be possible instantly. On these scales, the expected power of nanotechnology should fall somewhere between that of biology and the Star Trek transporter; between that of Robby the Robot and that of the Krell Machine. Such powers are God-like only if your imagination is limited, and your gods are of the slow and patient type. Still, they are impressive.

If nanotechnology should eventually be able to manufacture (or assemble) any reasonably small and cool object which can exist on a planetary surface, and do it on command, the next problem is who will be authorized to give the commands. Even if nanomachines are under docile control, their powers begin to resemble wizardry, and the way in which one may change the world with them (by speaking a word, or even thinking a thought) begins to look suspiciously like sorcery. Do we want that? Of course, in the virtual world *inside* a computer, it’s always been that way **[7]**. But the *Forbidden Planet* question is whether anyone, or any government, is safe in holding this kind of power over matter in the real physical world. With nanotechnology, we would get real "sorcery"-- but even with the best of intentions we might still find ourselves in the position of the sorcerer’s apprentice (think of Mickey Mouse in *Fantasia*). Even intelligent beings a good deal smarter than we are might not be wise enough to control such technology safely.

But this question, too, is shortly due to answer itself.

III. Mankind's Pending Ultimate Instrumentalities, Part B: The Computational Singularity

Unlike nanotechnology, the other main futuristic prediction of the 1980's regarding technology addresses a type of technical progress which is much easier to project, but (ironically) also evokes ultimate limitations which are much harder to imagine. The starting point for this second set of predictions involves the notion that information processing or "computation" can be done much faster than we do it. Further, there appear no obvious physical limits as to how fast computation may ultimately be done. Certainly, if there are limits, they are well beyond the power of our own low-powered and slow-switching brains.

Therefore, it must be possible to construct intelligences far superior to our own. Nor are the paths to doing this completely obscure, since in a real sense we already do it when we network, and allow many people to work on a given project too large for any single person to comprehend (a moon rocket or an economy). Or when humans use writing as an external memory aid, or work in concert with computers. Look about you. There is a reason why a modern city appears to be constructed by some designer smarter than any single person you've ever met. It literally has been. We're getting better at doing this, and this kind of thing will continue with a vengeance. As it does, it will assist in creating itself. This kind of progress in the speed of *progress itself* must inevitably lead to supra-exponential growth in information-processing or "thinking" ability.

Computing machines (first mechanical, then electronic) have been shrinking at an exponential rate for as long as we've been making them, and many people have sensed that there is something wildly empowering ahead. When the first kit to allow homebuilders and hobbyists to construct their own personal electronic computers was offered (late 1974), the device ended up being named the *Altair* (suggested by the 12 year-old daughter of the *Popular Electronics* publisher, after a *Star Trek* destination). The name somehow seems appropriate, for the Krell Machine is seen here, trying to be born.

Today, personal computer power has grown to levels quite unforeseen in 1974, and there is no end in sight. Instead, it seems that ahead is a kind of watershed-- or perhaps a waterfall. We are due to go over it. Such an event has been described in various terms for half a century, but we may refer to it as the *computational singularity*. The computational singularity corresponds to a singularity point in a mathematical function, where the value of the function approaches infinity (like **1/x** when **x** approaches zero). It is a time when total computational power rises to levels that are, if not infinite, at least qualitatively unimaginable. This is set to happen quite soon, if we continue at the present pace of advance.

Perhaps the first work of fiction to use this idea explicitly is the Vernor Vinge [VIN-jee] novel ***Marooned In Realtime*** (1986). In this tale, human time-travelers in time-stasis bubbles come out of suspension to find themselves on the other side of a curious rift in civilization, during which all humans have disappeared from the Earth, leaving the planet empty. No one who emerges from stasis understands what has happened to civilization, and since the travel is one-way, they cannot go back to find out. There are clues that the end hasn't been extermination. Possibly (Vinge hints) there has been an *Exodus* or *Ascendancy* or *Transcension* of some kind, since the computer technology of the civilization just before the rift has been clearly progressing exponentially toward a somewhat incomprehensible information-processing power. The implication is that mankind has perhaps “graduated” into some other kind of new mental life, much as happens in Arthur Clarke’s 1953 novel ***Childhood’s End*** (to which we will return -- Clarke’s fiction provides some of the first science fiction “mental millennium” genre stories, though the mental millennium in Clarke is not computer-generated).

Author Vinge, who in real life is an emeritus professor of computer science at San Diego State University, has also written formally in non-fiction about the concept of the "computational singularity" (1993, ref **[9]**). Vinge traces the idea at least as far back as speculations of J. von Neumann and S. Ulam, a pair of legendary figures who made deep marks in computer science, mathematics, physics, and complex systems theory in the 1950’s. Vinge also credits I. J. Good (1965, another polymath) with first pointing out explicitly that computer-design-of-computers leads to computer power progress which must be at least exponential. And indeed, here in the year 2002, already a year late, we don't yet have a **HAL 9000**, but we do already allow a great deal of chip design to be done by machine. We have no choice -- it's already beyond the capability of human designers.

The advent of true self-replicating nanotechnology, the first waterfall we discussed, may be difficult to predict. But recently there have been a number of suggestions that, by contrast, the *computational singularity* (which will be hereafter referred to simply as the "singularity") should be upon us within a generation or two. The reason for the more confident prediction is that information-processing power has been increasing smoothly and exponentially for a century, in a way which is much easier to extrapolate.

Roboticist Hans Moravec, in the classic future-shock robotics book ***Mind Children: The Future of Robot and Human Intelligence*** (1988) suggested that the unimaginable waterfall in this river of progress will happen about 2030 AD. Ray Kurzweil has recently updated and expanded Moravec's arguments in a book called ***The Age of Spiritual Machines*** (1999). In the book, Kurzweil suggests that during the last century, the doubling time of the figure-of-merit "computation power per dollar," which had been thought to have been relatively constant, has in fact decreased from three years toward one year. In other words, we used to have to wait three years to buy a computer twice as powerful for the same price, but with today's PCs, we now wait only 12 months for this to happen. So not only is the pace of change exponential, but the exponent itself is changing.

According to Kurzweil and others, the singularity is due not because of the sliding nature of the exponent (although this helps determine the time) but rather because of another key milestone: at some point in the process, our computers will become as computationally powerful as the human brain. This is projected to happen sometime between 2015 to 2030 AD, and the exponential effect insures that the personal computers 5 to 10 years later will be just as powerful. A few years *later*, it follows inexorably that computers as complex as the human brain will be mass-produced items, like digital watches or wind-up toys. Shortly after this happens, our computer networks are expected to suddenly (and nearly instantaneously from our perspective) get very, very *smart*.

Of course, a computer as powerful as the human brain does not guarantee the performance of a human-equivalent mind. Indeed, even humans themselves, if not programmed correctly, become less *Mowgli* than “wolf boy” – not much more than animals. One special thing about a human brain is its sheer connectionist capacity, and the ability to use this capacity to modify partly-inborn structural programs for learning. So many of the defining characteristics of modern humans are in their culture, not their bodies or brains; we are by now, in many respects, a *software species*. Similarly, the attainment of human and superhuman mental performance by computers depends on the ability to program computers heuristically by *experience*, in much the same way that we semi-program human minds today.

In such a scenario, simple learning programs become better learning programs until, at some point, they pass the Turing Test and become capable of some subset of human-level intellectual performance. The ancient Greek *sorites* paradox, as amplified by the philosopher Hegel, is then realized: an increase in mere (computational) *quantity* is mysteriously translated into a change in *quality*. We say that we now have a *system property,* or in modern parlance, an *emergent property*. In this case, the new property will be intelligent action.

That is the theory, but we are not without the beginnings of practice. Those who differ with the theory, holding instead that the human mind is a specially creative instrument in all circumstances, never to be duplicated, were dealt a severe blow in 1997 when the IBM computer *Deep Blue* defeated chess grandmaster Gary Kasparov. World champion Kasparov was thought by most chess experts at that time to have been as formidable as any player in chess history. Until he encountered *Deep Blue*, Kasparov had contended that the play of computers was typically rote-mechanical and unimaginative, in ways that a grandmaster could easily detect, and then exploit. Great chess was said to take imagination and creativity of a kind that would forever elude the machine. For a long time it pleased the vanity of humans to believe Kasparov, as he kept beating chess computers. Finally, however, came the day of reckoning, as an inexorable increase in raw computer processing power resulted in a self-learning chess-playing machine which (somewhat mysteriously) became capable of formidable chess imagination and insight. Even the programmers were not completely sure how it had happened.

Deep Blue now passed its version of the "Turing Test" for machine intelligence, for Kasparov felt for the first time that he was glimpsing a *mind* across the board from him. This may be the most interesting part of the episode, for Kasparov immediately accused the programmers of cheating, and of having a human chess master in contact with the computer during play. However, Kasparov was wrong. There was actually no one “home” within the programs that comprised the “mind” of *Deep Blue*. The programs which “creatively” dismantled and destroyed Kasparov’s strategies were running by themselves. Kasparov was indeed facing only a machine, not a human grandmaster, but now he *could not tell the difference***.** There is a lesson: this kind of thing can happen. And if it can happen here, it can happen in other areas of thought.

In the past, the field of Artificial Intelligence has suffered badly from making predictions that in retrospect could never have proven out in the time given. Even the supercomputers of today have brains only about as computationally powerful as those of insects, so they’ve really had no chance to think as well as humans do, no matter how well-programmed. Also it’s not very surprising that when given machine bodies, computers of today still interact with the world in somewhat insect-like ways. Indeed insects themselves often behave in many ways that seem to us to be somewhat stylized and mechanical.

Even with real insects, however, we see some of the principle we seek: a qualitative amplification of intelligence is possible, if we increase only total complexity. Hive-insect minds, working in a linked fashion, may develop the flexibility of much more complex and intelligent animals. A bee colony, for example, which has far more neurological processing power than any single bee, is *as a whole* capable of more complex learned behavior than are single bees. A colony will remember the location and times of flower openings, and is even capable of future-modeling or inductive behavior, rather like a vertebrate. If a dish of sugar-water near the hive is moved by a certain distance each day, bees eventually one day will be found clustering at the next projected or *anticipated* spot.

In the same way, we guess, things cannot fail to change qualitatively as electronic computers and their networks grow more complex. In the future, as these networks become more capable, they will presumably mimic brains that are further along in evolutionary scale of complexity. Today's insectoid machines will one day act like lower mammals, then higher ones (toy makers are already busily modeling the behavior of dogs and babies with 8-bit microprocessors, and doing surprising well at it). We can guess that along the way, machines will pass more and more Turing Tests, in which their behavior cannot be told from that of a human, over ever-wider areas of human "expertise."

Again, in making such projections, we run up against the past bad predictions of Artificial Intelligence enthusiasts. A.I. has always seemed forever in the future. But we should be careful of such things. The moon landing, gene therapy, and mammalian cloning were old science fiction ideas that seemed forever in the future, too, but they didn't stay there. Eventually, if computers continue on their present path, Artificial Intelligence, too, will come. Then we will presumably have robots like HAL or Robby, who answer questions in a flexible and non-mechanical way. (Complimented on the nice high oxygen content of the Altair IV atmosphere by humans making small-talk, Robby comments dryly: "I rarely use it myself. It promotes rust."). At that point, we'll have to begin worrying about whether or not such devices are not the equivalent of animals, or perhaps are even something more.

There has been argument here too, of course. Vinge himself has remarked **[8]** that the super-accelerated mind of a dog (say) would still not be human. But we may note that dogs as we have known them are particularly crippled by a short attention span and a relatively poor memory, neither of which would be expected problems for a computer-enhanced dog-mind. Indeed, Vinge himself has recently written some excellent science fiction discussing the value of having monomaniacal attention-span at one's command, if only one can also leave some executive functions in control of it **[10].** A dog is also notably crippled by lack of hands and by lack of brain circuitry which allows rapid recognition, identification, and use of sounds and visual symbols which make up language (chimps have some of this). Add all these things, plus some mental quickness and some training and teaching, and it seems likely that a dog will no longer be a dog. Just what it *will* become, given enough time and experience, is an open question **[11].**

If we assume that self-programming ability follows processing power, very soon after the point that computers of human brainpower are mass-production items, we may expect that computers will attain the total information processing power of all human minds on the planet. They will have long since become the experts in the design of more complex computers, just as they are today the reigning experts at chess strategy. At some point not long after that, computers will recapitulate human history, human culture, and human thought. They will then teach each other everything we humans know in a matter of years (months? days? hours?), and then move on. The whole thing will happen in a flash, and if it happens at all, will certainly happen long before we're really ready for it. The "flash" seems inevitable before the end of this century, and seems quite probable (given even modest extrapolation) before the middle of it. And, of course, we'll be unable to stop it, anymore than we can stop anything on the Internet. Before we know it, it will be done.

In theory, either full nanotechnology *or* the computational singularity might happen before the other. But whichever arrives first, it seems probable that the other will then immediately follow in consequence. Nanotechnology, after all, requires molecular-scale self-replicating computers, and such machines should rapidly be able to grow and wire themselves in three dimensions to the complexities needed for the singularity to occur. In a similar fashion, an evolved computer which is far faster and brighter than we are, will soon figure out how to manipulate matter on the atomic scale with self-replicators, and will then do so in service of other goals, unless actively prevented. Thus, nanotechnology, whether it arrives first not, seems destined to be the incarnate "muscle" of the singularity Artificial Intelligence.

One might imagine optimistically that we might prevent such a connection, with safeguards which prevent super-intelligences from interacting with the physical world, except perhaps by something like censored E-mail. On second thought, however, any careful isolation program may be doomed. Just as well to expect a bunch of chimpanzee guards to keep humans from escaping from Alcatraz. If a super-intelligent computer has enough contact with the world to be very useful, it will probably have enough contact to subvert some of its captors into aiding it to escape.

An Artificial Intelligence might amass wealth, for example, and with that wealth influence the passage of laws in democracies. It might also simply bribe outlaw humans and outlaw governments. People who imagine that governments can control super-intelligent computers might consider just how much control governments today have over junk-E-mail, the Internet, or very large multinational corporations. Self-aware computers (which will be running the more successful corporations by that time) will be far faster and more slippery than anything we've dealt with thus far.

IV. Penalties For Playing God or Wanting To

After such an escape of Artificial Intelligence or nanotechnology into the "real world" and private hands, then what? Mankind does not have a good record for handling destructive technologies. We have avoided global exchange of nuclear weapons till now only by a hair's breadth, and would not have come this far if *all* governments had nuclear weapons, and still less if all *people* did. Coming soon now, however, is something as pervasive as the personal computer and cell phone, but with the power of mass destruction too.

There is the problem of deliberate "bio" or "nano" warfare. Viruses and bacteria as we know them are already much like assemblers, and can be made worse (for example, imagine HIV with its present latency period, but with the infectivity of influenza). There is also the problem of natural replication mutation accidents, which correspond with the emergence of new wild viruses, like Ebola, HIV, or even the latest strain for the flu. As in any self-replicating system, parasitical forms may emerge in nanotech systems. An uncontrolled self-replication/assembler system can be imagined. It popularly manifests itself in the prediction-genre as a creeping, corrosive *gray-goo*, a kind of undifferentiated assembler-cancer. Such stuff causes disaster, because like some super-corrosive bacteria or slime mold, it exists merely to transmute anything it touches into more of itself. Some say the world will end in fire, some say in ice (as the poet Robert Frost writes). Now, there is a third and more insipid option: perhaps it will all just melt into corrosive amoeboid sludge **[8].**

Those who favor fire may note that easy manufacture of nuclear weapons by uranium isotope separation should be a fairly straightforward subset of self-replicative manufacturing technology; yet no foreseeable technology, including nanotechnology, can provide a defense against such weapons. So there are many ways in which the coming world will get scarier **[12]**.

Very well -- perhaps we have to "Let go" and "Let God" (as a bumper sticker says). Perhaps the advanced machines will end up doing everything for us, and in true *Deus Ex Machina* style, everything will be fixed-up, and come out all right in the end. We like such endings. Culturally, the relative closeness of the singularity has visited on its truest believers much the same effect as belief in the imminence of The Second Coming. The complex set of apocalyptic ideas which parasitizes and sometimes immobilizes adherents to certain brands of Christianity, now in other guises seems to handicap certain alarmists and "cybernetic totalists" (to use Jaron Lanier's phrase) with visions of Technological Salvation, or Techno-transcendentalism. First it was Cryonics, then Nanotechnology, and now Singularity (all capitalized as religions, or at least political affiliations) which will get us to the "End of Time." And all perhaps without the conventional God. All of these ideas can serve as an apocalyptic religion, if conveniently simplified and the most scary parts are left out. We are promised the apotheosis of mankind.

At least the techno-evangelicals don't wear placards saying "**THE END IS NEAR / REPENT NOW!**" Actually, there doesn't seem anything much to *do* in the Religion of Singularity except spread the Good News (hence, perhaps, this essay). And, of course, one must believe. To be sure, there exist some who do seek to bring a more critical eye to the whole idea-set. The reader is referred to www.SingularityWatch.com **[13]**. Still, the whole thing does cause a certain amount of unease.

It's easy to place the sources of that. To begin, what will be the nature of these coming A.I.

super-intelligences? Will they be nice, or will we get, instead of *Forbidden Planet,* perhaps *The Forbin Project*? Or *Terminator*'s Skynet? Is there nothing else to do in the way of safeguards?

In *Forbidden Planet*, Morbius' powerful robot servant *Robby* has been explicitly constrained by Morbius to observe Isaac Asimov's "**Three Laws of Robotics**" [**Editor's Note:** The Three Laws of Robotics are as follows: (1) a robot shall not harm humans; (2) a robot shall follow human orders except in the case where such orders would conflict with the first law; and finally (3) a robot shall seek to preserve itself, except in such case where its actions would conflict with either the first or the second laws]. The Krell Machine, by contrast, is an infinitely dangerous servant precisely because it has not been preprogrammed with Asimov's Three Laws in mind, and the Krell have evidently made a monumental error on this point.

We would like to take a precautionary lesson from the noble Krell. Could we perhaps hardwire Asimov's Three Laws permanently into machines that are smarter than we are? Alas, it may be that the answer is no for machines that "rewire" themselves, which is what they will have to be capable of, if they ever *are* to become smarter than we are. Here is the rub of A.I.: we cannot directly program minds to be better than ours, because we don't know how; and yet if they program themselves through learning, we won't then fully understand them, and certainly won't then be able to perfectly control them and predict their behavior. There is no such thing as immutable "hardwiring" when software is in control. Anything created by evolution may be *un*created, or gotten around by a similar process (as Asimov himself pointed out in later life, on thinking about the future of robotics).

In creating super-intelligent robots, then, we can only face the key problem of every responsible parent, and place our hope in the Hebraic injunction: “Train up a child in the way he should go, and when he is old he will not depart from it.” Or will not depart too badly, we hope.

And what about the other Krell lesson? Leaving aside what the computers may want, what about what *we* desire from the genie? What if the fates punish mankind by giving it what it wants, on both conscious and unconscious levels? Our experience with children and animals, not to say ourselves, makes us suspicious (to say the least) of what occurs then. The effects of our present fad and impulse-driven market economy (not that the author sees better alternatives) on ourselves and the biosphere are frightening enough. What happens when these effects and externalities all become infinitely amplified via technical means?

According to our cultural mythology, both before and after the advent of science-fiction literature, poets have classically laid heavy penalties on those humans who sought to steal knowledge from the Gods. The penalty is ostracism and worse: (1) Prometheus was chained to a lonely rock and tortured; (2) Adam and Eve, according to *Genesis*, were punished for their sin of disobedience, being evicted from The Garden of Eden and sent into an uncharted Earth. This prevented them from subsequently eating of "The Tree of Life" and achieving immortality (becoming Gods).

Science fiction as we moderns recognize it, properly began (1818) with the novel ***Frankenstein*** (subtitle: ***Or: The Modern Prometheus***), in which the monster, as a price for its unnatural science-given life, is cast out of society to wander -- forever looking through the window at the celebration, forever seeking one of its own kind to talk to or love. The monster suffers the social tortures of adolescence, and author Mary Shelley, we may not be surprised to learn, was a motherless child who wrote the book while herself still a teenager.

Shelley, in her later writing, sought other expressions of alienation: one of her works (***The Last Man*****,** 1826) features a man who is all alone on a completely depopulated Earth. Since Shelley, the ruined or deserted Planet, from Nuclear Winter to Silent Spring, has naturally come to be associated with visions of higher technologies and the far future (see H.G. Wells' ***The Time Machine*****,** 1898). The Krell Machine in its many forms typically inhabits empty worlds (just as Prospero, in ***The Tempest***, inhabits a nearly deserted island). Krell Machines of various kinds sit unused and lonely in the ruins of lonely cities on the edges of forever -- their former users having either been destroyed or else left for their dreams, leaving the shards and husks of more mundane realities behind.

Perhaps the image of "wasteland containing a doorway" is a fictional metaphor, which arises from our childhood experiences of being lost outside the home in a world we do not understand. Certainly it makes for a better story to be faced with a functional alien artifact that has no User's Manual. Larry Niven's early short story *Wrong Way Street* (1965) and Frederik Pohl's ***Gateway*** novels (1977-) contain an entertaining use of this plot device: long vanished aliens have left a deserted space port, and some of the semi-automatic spacecraft still work **[14]**. Push the button and you go to wherever that ship is programmed to go (now, which of these thingamabobs do you suppose is the *fuel gauge*..?). Such mysteries are always dangerous, and they are not always resolved. In Algis Budrys' novel ***Rogue Moon*** (1960) humans use disposable duplicates of themselves to explore a large and still-working maze-like alien machine found on the moon. The moon artifact kills people who explore it in various gruesome ways, apparently as a side effect of a true design function which humans never do figure out.

It does seem to be a nearly universal idea in science fiction that the result of attaining ultimate technological power must be that those who have access to it vanish like 16 year-old boys with the car keys. We don't always know where they go, but their disappearance is expected. Stephen Spielberg's film *A.I.: Artificial Intelligence* (2001*)* typifies a now-standard mystery form. *A.I.* is a straightforward re-telling of the **Frankenstein** story, with all of its sub-texts of social isolation, child-abuse, and creators who fail to live up to their responsibility. The protagonist, an artificial child, is abandoned like an unwanted pet to wander the Earth as an outcast, and finally is put out of his misery by being accidentally cryo-preserved (Shelley‘s original **Frankenstein** also begins and ends in the arctic, as a metaphor for isolation and loneliness). When the robot child wakes, humans have vanished, the cities are in ruins, and the child is surrounded by alien mechanoids whom he still asks pitifully for his human mommy **[15].** That's meant to give you the creeps, and indeed it does. *A.I.* did not do as well at the theaters as it might have, possibly because, like the robot-child himself, the film jerks too many human emotional strings, and does so too vigorously and too artificially.

We frequently do not know where civilizations go when they hit the singularity in fiction, but sometimes they leave behind deliberately-cryptic messages. For example, in Robert Forward's early treatment of the idea (***Dragon's Egg*****,** 1980; ***Starquake*****,** 1985), the alien action is set on the surface of a neutron star. The indigenous intelligent life is somewhat like electronic computers, inasmuch as their nucleonic brain "chemistry" allows them to think a million times faster than humans can [16]. In these novels, humans initially arrive in orbit around the neutron star to discover the inhabitants in a very primitive state. Humans cannot visit the star's surface due to its fantastically high gravity, but communication by flashing light is eventually established. As the neutron-star creatures are taught by humans, however, they rapidly assimilate human culture, and just as rapidly, surpass it. Then, suddenly, to the surprise of the starship crew, the world below them is empty. The aliens have reached their own "singularity" and (of course) disappeared. They leave behind nothing, save for a few condescending clues-- the litter of "Ascended Beings" who now don't wish to interact with primitive humans, until we are ready. This occurs in a novel published a year before ***Marooned In Realtime***, so the idea was current in certain circles by then (Vinge, for one, had been talking it around for a few years).

The ultimate humiliation may be an empty world containing vestiges of advanced beings who *could* talk to us if they wanted to, but don't seem to *want* to [17]. We've seen a similar theme in *Forbidden Planet.* The superhumanly intelligent Dr. Morbius is a creator beyond good and evil, and doesn't at first want to communicate with ordinary men. There is something of Nietzsche about him (why else is he a *philologist*?) He has come to identify with the super-human. The human I.Q. does not impress him, for his own brain has been augmented by the Krell Machine, which is an intelligence-enhancer as well as a physical-realizor of ideas.

Morbius' technology and his intelligence are in the realm of magic, a la **Clarke's Law**, and at the end of the film, Morbius wears the wizard robes of Shakespeare's Prospero to illustrate this. We are fascinated that, like Prospero, Morbius has difficulty escaping his own animal passions. As even a much more advanced species on Altair IV could not. Or so the humans in the film are led to presume.

In *Star Trek*'s most light-hearted invocation of the Krell Machine (Theodore Sturgeon's *Shore Leave*, 1966) the crew of the *Enterprise* beam down to an apparently empty planet, only to find that it too hides machinery which has the job of making fantasies into realities. After being harassed by the incarnate results of their idle thoughts, the crew finally encounters the planet's alien Owners. The Owners use the technology for recreation (and for medical care — they repair a "dead" Dr. McCoy as easily as any machine). But they tell Captain Kirk that they (the Owners) are too advanced to meet humans. Now run along and play. But thanks for asking **[18].**

As with the scenario of nuclear war, it is traditional for planets to come out of the other side of the singularity depopulated, or worse. Science fiction is full of cautionary wastelands and ruins, markers of a time when humans stole Promethean fire and were burned in it. Authors of science fiction, for their part, write *past* the singularity simply because it's nearly impossible to write convincingly *into* it and keep a good and readable story with characters which we can care about and identify with. It's too strange. But there are many "fly-bys" of such apocalypses in the genre.

Childhood's End (1953), the Clarke novel mentioned earlier, contains one. If *alienation as the price of technical advance* is the primal theme of all science fiction **[19]** then it can be added that Arthur C. Clarke's story plots (in particular) often involve alienation with some continued and distant communication. Clarke's characters are often beyond help, but they can always still talk while they are trapped, or while meeting their seemingly inevitable doom. In ***Childhood's End*** the role of the outcast monster is played by alien creatures called the "Overlord*s*." The Overlords are inhumanly intelligent and ethical but physically unlovely beings who are destined never to be able to make the evolutionary leap to higher consciousness, and who must therefore spend eternity on the outside of the party, looking in. They are alienated aliens; monsters who are troubled with their own monsters.

At the end of the novel, the Last Man on Earth stays to fatally witness mankind's transition to higher being. He continues to talk by radio through the last minutes of his life to the retreating Overlords, as the Earth itself begins to become transparent, in a scene which reminds us once again of Altair IV, the wizard Prospero, and some of the more famous lines from the play that was the inspiration for *Forbidden Planet*:

Our revels now are ended: these our actors,
As I foretold you, were all spirits, and
Are melted into air, into thin air:
And, like the baseless fabric of this vision
The cloud-capp'd towers, the gorgeous palaces,
The solemn temples, the great globe itself,
Yea, all which it inherit, shall dissolve,
And, like this insubstantial pageant faded,
Leave not a rack behind: we are such stuff
As dreams are made of, and our little life
Is rounded with a sleep...

And this is all we can really say, as Earth or Altair IV disappear in the aft-viewplate of our imaginations. The problem with the singularity is that there is apparently no way to "survive" it (pace the tongue-in-cheek Vinge sub-title *How to Survive in the Post-Human Era* **[9]**) because it is the nature of the singularity to change beyond all recognition even the basic concepts of humanity, life, individual identity, and survival.

Particularly "individual" survival. A central problem in our imagination of what the singularity might be like, is that the interfacing of brains and computers in the singularity must result in a vicious melding of various kinds of minds. Vinge remarks that "[a] central feature of strongly superhuman entities will likely be their ability to communicate at variable bandwidths..." This is a safe and nearly tautological prediction, for breadth of bandwidth is all that defines whether "communication," as we usually understand the word, is taking place at all. *Communication* is generally not a word we use in connection with the mind's *internal* affairs. “Communication” therefore requires two or more minds --- yet if bandwidth is too high, individual minds must disappear, and only one group-mind is left. Thus, within a grouped computational being, minds and sub-minds are *defined* only by bandwidth. Imagine being "you" only when you close the door on the party, or they close the door on you. If the door is opened wide, however, “you” cease to exist, and you and they become part of a Larger You (or collective *Us*) **[20]**.

Such Borg-like problems plague our predictions. So much so, that writers considering the very far future have usually had to split some powers of technology off, in order to have any recognizable human culture to deal with at all. For instance, Frank Herbert, in his ***Dune*** series, simply outlaws machine intelligence. Too much telepathy combined with too much technology makes it difficult to generate recognizable dramatic tension, which comes from recognizable *characters* with *problems* we primitives can care about.

We let one more empty-planet novel serve as a final example. Arthur C. Clarke's novel ***The City and the Stars*** (1956, contemporaneous with *Forbidden Planet*) deserves mention in anticipating many ultimate technologies. This novel is set a billion years in the future, in a utopian metropolis called "Diaspar." Diaspar's machinery can manufacture anything on demand, including human beings. Indeed, the city's very inhabitants are a random collection of people from the much greater store available in the city's memory banks, something like books circulating from a central library. Each inhabitant lives a thousand years, but also recovers his old memories from previous incarnations, giving him functional immortality. And yet, the novel's main character, restless to explore, eventually escapes his version of the Krell Machine.

Outside Diaspar, he finds the traditionally empty Earth, uninhabited except by a few mentally-advanced communities of humans. These people deliberately eschew technology, and live a rural, somewhat Amish-like existence, complete with normal human reproduction, normal aging, and standard death. Significantly, however, they are telepathic; and thus experience a sense of community and communal immortality, which they find to be a satisfying replacement for technological immortality.

Thus, Clarke's immortal Diasparians pay for their technical utopia with severe communications and social isolation problems, and with no way to satisfy the urge to explore. It is difficult to imagine the kind of life-style that would result if they were not thus crippled.

Yet the sum of the gifts of both Clarke's alternative worlds is exactly what we must contemplate for ourselves -- not a billion years from now, but very possibly in the next century.

The name "singularity" to describe such a state-of-being is appropriate because, as is the case with a black hole, the singularity looks different depending on whether it is viewed from outside, or from the point of view of an observer falling into it. We have readable fictional scenarios only for the outside. For all we know, however, perhaps *these* outside views are the futures that will ultimately come to pass for mankind. After all, it is by no means certain that mankind will either be destroyed *or* entirely uploaded/assimilated into something non-understandable. There is a third possibility: mankind might be left in the dust like those old computers (or toys) in your garage that you're never going to play with again (Spielberg and Aldiss, in the film A.I., work this "*Puff, The Magic Dragon"* theme masterfully **[19]**). If the singularity had been called the "Techno-Rapture" it should also be remembered that a fundamental feature of the Rapture is that some go, while some are left behind.

Will those who wish to go into the singularity, have a path to do so? One of the key issues determining what kind of future we get may be the timing of the development of a full brain/computer interface. Whereas computers may be made to talk to each other with relative ease, the human brain is not wired to accept or process input more complex than sensory data. Indeed, in *Forbidden Planet*, all but a few human brains overload and burn out when exposed to connection with the Krell technology **[21]**. It will not be a trivial undertaking to directly connect brains with computers or to technologically connect brains with each other (mechanical telepathy). Virtual reality is technically simple compared with, say, constructing a system in which one can sort through and "remember" items in a computer database as easily as sorting through one's own memories. Thus, it may be that the planetary web of computers systems will exceed the sum of human intelligence well before the interface problem is solved. If events happen in this order, it will be up to the Artificial Intelligence, not mankind, to figure out how to put the full link between machines and humans safely into place. There is no guarantee that the singularity A.I. will choose to do so.

There are dark possibilities at this point. Perhaps the Artificial Intelligence will simply protect itself and impatiently go on, without us. Perhaps (worse) it will even leave humanity with some kind of technological lock, in order to prevent development of the computational power necessary for such uncouth creatures as ourselves to follow. Singularity-struck societies which leave any beings "behind" may even represent a kind of threat to the ascended beings who have gone before. The reason is that such stuttering "techno-adolescent" societies can be expected to attain new technical singularities regularly. With each one, they would unleash new species of Ascended Intelligences into the company of those who have gone before. Might some of these new emergents be pathological? The jury is out — it is too early to guess. But if so, such societies might therefore be under careful watch by those who have already transcended. They may, conceivably, even be under quarantine.

“What?" you say. "Surely these machines will let mankind 'upload' or mind-link with them, and join the party **[22]**. Won’t they? They have to!”

Er....don’t they?

If not, we can glimpse *that* future — it's the main one we are familiar with from science fiction. And, likely, also familiar with, from some of our own early adolescent experiences of being shut out of the world of adults. We know what things will look like then. They will look like being locked out by an intelligent computer ("Open the Pod Bay Door, HAL!") who not only controls our technology, but also tells us that conversation can serve no further useful purpose **[20, 23]**. Mankind would then forever be the chained Prometheus, forever the orphaned and lonely Frankenstein’s monster looking though the window -- the subject of the ultimate snub. Indeed, we would be forever Caliban, left alone on an island-Earth with the wizards gone -- and not even comforted by the whisperings of spirits that have long since been freed.

NOTES:

[1] E-mail address: sbharris@ix.netcom.com. The author appreciates any constructive feedback.

[2] In Arthur C. Clarke's ***2010: Odyssey Two*** (1982), self-replicating all-purpose monolith machines, the alien Krell Machines of this tale and its successors, turn Jupiter into a small star. The humans in Jovian orbit get away just in time.

[3] Feynman's original 1959 talk, later published in CalTech's *Engineering & Science* (February 1960) is available at http://www.zyvex.com/nanotech/feynman.html .

[4] http://www.zurich.ibm.com/news/96/n-19960112-01.html. Drexler first published in the peer reviewed journals on molecular manufacturing in 1981. Readers interested in the history and current progress in nanotechnology, including most issues discussed in this essay, are referred to http://www.foresight.org.

[5] A serious attempt to predict what such future medicine would be like is Robert A. Freitas, Jr.'s **NANOMEDICINE** (Vol 1), Landes Bioscience, 1999.

[6] A human being is not the atoms making him up, anymore than a *novel*, an insubstantial thing, is the atoms making up a particular physical book or audio tape. Atoms in the body are replaced in metabolism, but the person remains. In theory, all atoms could be completely replaced, and yet the *person* would still remain, as a *pattern*. A human being is information,

not matter. Such information can theoretically be extracted on a molecular scale, sent from here to there, and reconstituted as a pattern in new matter.

To make an "effectively identical" duplicate of a person, such a process doesn't have to be done for each individual atom in a body, because most positions of most atoms in a person don't make any differences that we care about. For example, protein molecules and cell organelles can be produced as generic copies of a single design, once identified by position (a person might have less than 70,000 different protein/gene designs, and much of the rest of his "protein" information is where each copy is, and how it has or hasn't been modified in place). On a larger scale, many cells and even tissues can be generically specified the same way — for example, you probably don't care if all the glomeruli in your kidneys are replaced by many exact copies of a few of your best-performing ones. The important information in transmitting a human being will be in the connections of his or her neurons, and information regarding the delicate modification of proteins in the synapses. These form memories, some of which are not shared by any other human, and are thus irreplaceable. Some parts of a copy count more than others, if you care about performance. For example if we want a duplicate player piano to play a recognizable piece of music, we must be particularly careful about the position of the holes in the new piano scroll, but may be less careful about things like what the keys and pedals are made of, how the piano is painted, etc.

[7] As Vernor Vinge was also among the first to point out, in his short story *True Names* (1981). For the 20 years since this work, several sub-genres (see for example W. Gibson's ***Neuromancer*** 1984), have explored the ways in which power inside a computer network may give power in the external world. The recent film *Matrix* (1999) is a descendant of this tradition, highlighting ways in which programming power and physical power will meld in the future. See also **[22]** below.

[8] If this happens, all is not quite lost. These is a minor consolation in that one suspects that gray-goo will be subject to the same evolutionary pressures as the rest of life, and that (even if it arises) it won't stay primitive forever. See [4] for thoughts on grey-goo defense.

[9] Vinge's essay is available at:
http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.ht ml

[10] See Vinge’s novel ***A Deepness in the Sky*** (1999), another novel in which humans wait above an alien planet, patiently teaching, until the culture below progress to equal that of the space-farers. Vinge's chief horror-source in this work—the idea of finding yourself with full intelligence but slave to the grip of a monomaniacal madness, goes back in literature at least to Edgar Allen Poe's 1835 short story *Berenice*
(http://bau2.uibk.ac.at/sg/poe/works/ber enice.html).

Interestingly this particular Vinge novel does not posit singularities when civilizations grow sufficiently complex, but rather suggests inevitable breakdowns involving bottlenecks in communication within civilizations, leading to collapse and barbarism, much like Asimov's Foundation series (see the history of the Roman Empire).

[11] I suspect that such augmented animals, even if never capable of formal operations, may yet advance far into progressive academic political thought.

[12] Uranium isotope separation is more a physical than a "chemical" process, but it is still amenable to processes which could be performed on a small scale, then duplicated into practicality by a self-replicating manufacturing capability. The special problem with nuclear weapons is that they generate temperatures of tens of millions of degrees, and therefore no imagined material can stand up to them. For gray-goo or bio-warfare weapons or accidents, there is always a possible nanotechnological defense (in the literature, police nanomachines are naturally known as blue-goo). However, a defense against actual nuclear weapons (if you don't count distance!) falls into the realm of techno-fantasy. Such a defense joins science fiction ideas like faster-than-light travel and backward time-travel, as a technology that would require new physics, or new kinds of matter, and which may therefore *never* come to pass. This is in sharp contrast to rest of the engineering developments discussed in this essay, which require mere technical progress, but no new physics.

Under threats of various kinds of mass destruction in the hands of individuals, many pre-emptive defenses will be tried. Partly due to security concerns, it is another inevitability of the future that, shortly, none of us will have much privacy. People who have lived through the last 30 years have already noticed that the increasingly computerized world is rapidly developing a certain “lack of slack,” as information regarding anything you’ve ever done which created a record anywhere, threatens to become almost instantly available to nearly anyone who has money to pay for it. Many public places are now under continuous video surveillance, and very soon, they all will be. With computer visual image-recognition, soon the power will be available to track your travel, and all your public activities, just as we now track 18-wheel trucks on the highway. It's all a matter of processing power, which (as we have seen) discounts at 50 percent a year, year after year. If it's expensive to keep tabs on you now, it will be half as hard next year, a quarter as hard the year after that, and so on. Efforts to stop it will subjected to far more resistive economic pressures than efforts to stop junk-mail and junk E-mail, and we've seen how effective trying to do that has been.

[13] See http://www.singinst.org/ for an unabashedly boosterish singularity-promoting site. *Singularity Watch*, another organization, has been attempting to develop an "Academic Conference on Accelerating Change" by getting multidisciplinary scholars to more objectively evaluate the quality of evidence for "technical acceleration" of the kind that feeds on itself. My particular thanks to John Smart, organizer of the www.SingularityWatch.com site, for many helpful comments on this essay.

[14] The 1965 Niven story is notable for describing alien technology which is able to grow crystals of any type and size "atom by atom" from basic building materials. Again, this is the vision of Robby the Robot. But Niven thinks bigger—he describes *rocket motors* thus made from single diamond crystals--- as it happens, the exact image of techno-wealth which will figure prominently in the popular work of K. Eric Drexler, a generation later (see [23] below). Unlimited rockets and gems: the message is that nanotechnology has something for everyone; for *him* and for *her*.

[15] Stanley Kubrick, in true *2001: A Space Odyssey* style, has given us an ending which is rather ambiguous and frustrating, unless one knows something of the original script conceptions. For these, see
http://www.visual-memory.co.uk/faq/index2.html. The creatures at the end of the film are meant to be advanced earth robots, not aliens. The problem is that they know so little of their own origins that they may as well *be* aliens, and they essentially function in the plot as such.

[16] Hans Moravec has suggested that entire neutron stars (nicknamed "neuron stars" by Damien Broderick) might be turned into the kinds of brains that Forward describes. However, the interiors of such stars are uninteresting crystalline neutronium (perfectly identical neutrons jammed into one another like sardines, forbidden by conditions to be anything else); such a bland system is not complex enough to support information-processing. Neutron star surfaces are likewise not especially interesting, because pressures there require matter to be fairly normal (this looks bad for Foward's surface creatures, who have no way to pressurize their brains into nucleonic chemistry). Only the very thin subsoil crust of a neutron star has the complex mix of nuclei transmuting to other nuclei (dripping neutrons and gamma rays while doing so) which seems complex enough to support computation. This may be close to the ultimate computer, since each 1-centimeter layer of neutron star "crust" has a total mass approximately that of Earth, but compressed into a shell only 30 kilometers in diameter.

To be sure, more massive and extended planetary-sized computers are possible in theory (see Broderick [23],) but signal transmission lag-time for these much larger objects makes them much less attractive for the purpose.

[17] Classic stories of computers which become superintelligent, then simply uncommunicative, are Stanislaw Lem's *Golum XIV* (1981), and Larry Niven's *The Schumann Computer* (1979).

[18] In a blacker *Star Trek* episode of the same year (*What Are Little Girls Made Of*), the role of the Krell Machine in the ruins of the empty planet is played by a robot-making device which can either upload or mirror humans into mechanical bodies. It is attended by the usual mad archeologist (this time a replicant), a pretty girl (this time a robot), and finally an ancient alien robot "servant" who finally remembers the mental formula by which the robots once found a way around the safeguards of their makers, and managed to destroy them.

[19] Brian Aldiss suggests only that the central theme of science fiction is *alienation*, but the connection of alienation with technology is certainly implied and understood. See Aldiss' excellent science fiction review **The Trillion Year Spree** (with David Wingrove, 1986). Aldiss also happens to be the author of the short story ***Supertoys Last All Summer Long*** (1969), upon which Kubrick/Spielberg's A.I. film is loosely based. In the movie *Toy Story* (1995) we experienced the dramatic tension of intelligent toys (*beings*) being treated as mere toys (i.e., as *things*, not people). Aldiss and the movie *A.I.* work this theme even more explicitly, since the android-makers of the film, now in the role of Dr. Victor Frankenstein, are fully aware of what they are doing. We have also memorably seen this in Ridley Scott's film *Blade Runner* (1982).

[20] See Alfred Bester's ***The Demolished Man*** (1953) for one of the earliest and best views of a fully telepathic society. Individuation will be something of an act of will in such circumstances. Although we cannot predict what life will be like on the other side of the singularity, we may guess that social strife in the style of "who's not talking to whom" may long survive problems of physical want, or even problems of mortality, in the future.

It is worth noting that, so long as our present notions of physical law hold, there will still always be circumstances in the future where physics dictates no choice in these matters. The physical size and mass (self-gravity) of any computer structure eventually must limit the maximal complexity of the computer, and on these distance-scales the speed of light must limit the bandwidth of two-way interactive communication *between* maximally large and complex computers (minds). In the future, it may be comforting to know that the day of the individual will never completely pass, since some kind of individuation on the fastest time-scales seems destined always to be enforced by communications delays. Arthur C. Clarke, Brian Aldiss, and Vernor Vinge have all written fiction in which this is an explicit sub-theme.

[21] Brain burnout from brain-boosting connections is common in science fiction—for other examples see Piers Anthony's ***Macroscope*** (1969) and Vernor Vinge's ***A Fire Upon the Deep*** (1992). The Vinge novel is particularly interesting in treating several cases of individuation forced on group minds by communications problems, as discussed in the previous note. (In this novel also, intellectually transcendent beings last as "Gods" for only a decade or so before they become incommunicative, and disappear.)

[22] People who are tired of the ills and emotions of the flesh may wish to simply transfer their consciousness to mechanical bodies and be done with it, as Moravec suggests seriously in ***Mind Children***. See William Butler Yeats' *Sailing to Byzantium* (1928) for an early romanticized view of this option. An especially creative cyber-existence science fiction tale, in which a man's consciousness is uploaded into an animal and finally a computer-world in which he can have his every fantasy, is John Varley's *Overdrawn at the Memory Bank* (1976). For an excellent book-length fictional treatment of this theme, see Charles Platt's ***The Silicon Man*** (1991). These tales explore one type of scenario in which human consciousness is mechanically separated from human flesh. They do not treat the far more complex situation (because there would be no understandable story if they did) of what may be expected to happen when human and "machine" consciousness become intermingled and interconnected to any extent desired, and when manufacturing capability makes the distinction between synthetic and biological "bodies" no longer meaningful either.

[23] For a delightful romp through many of the possibilities discussed in this essay and more, the author suggests Damien Broderick's book-length treatment of the singularity, titled ***The Spike*** (Tor, NY, 2001). Broderick is a long-time futurist (who used the term “virtual reality” in our modern sense in a story as early as 1976, and has been credited for inventing it), and who is up to date on ideas about what’s coming, and the history of these ideas (see http://www.panterraweb.com/the_spike.htm). Broderick points out that engineer Theodore B. Taylor first called self-replicating von Neumann devices "Santa Claus Machines" in a 1978 essay, in the same sense that I've referred to them as Krell machines. In this essay, Taylor discusses such synthesizer-devices which can make anything on demand-- as Robby the Robot can-- but the immediate illustrative application Taylor has in mind is the use of such replicative devices to mine the moon. This is possibly the entry point to the idea for the (then) L5 space-colony enthusiast K. Eric Drexler, who would start writing of his own about miniature Santa Claus Machines, just three years later.