



The Coming Technological Singularity

Vinge, Vernor

Published: 1993

Categorie(s): Non-Fiction

Source: Feedbooks

About Vinge:

Vernor Steffen Vinge (born October 2, 1944 in Waukesha, Wisconsin, U.S.) is a retired San Diego State University Professor of Mathematics, computer scientist, and science fiction author. He is best known for his Hugo Award-winning novels *A Fire Upon the Deep* (1992), *A Deepness in the Sky* (1999) and *Rainbows End* (2006), his Hugo Award-winning novellas *Fast Times at Fairmont High* (2002) and *The Cookie Monster* (2004), as well as for his 1993 essay "The Coming Technological Singularity", in which he argues that exponential growth in technology will reach a point beyond which we cannot even speculate about the consequences. Source: Wikipedia

Copyright: Please read the legal notice included in this e-book and/or check the copyright status in your country.

Note: This book is brought to you by Feedbooks

<http://www.feedbooks.com>

Strictly for personal use, do not use this file for commercial purposes.

Electronic License

(c) 1993 by Vernor Vinge (This article may be reproduced for noncommercial purposes if it is copied in its entirety, including this notice.)

The original version of this article was presented at the VISION-21 Symposium sponsored by NASA Lewis Research Center and the Ohio Aerospace Institute, March 30-31, 1993. A slightly changed version appeared in the Winter 1993 issue of Whole Earth Review.

Abstract

Within thirty years, we will have the technological means to create superhuman intelligence. Shortly after, the human era will be ended.

Is such progress avoidable? If not to be avoided, can events be guided so that we may survive? These questions are investigated. Some possible answers (and some further dangers) are presented.

What is The Singularity?

The acceleration of technological progress has been the central feature of this century. I argue in this paper that we are on the edge of change comparable to the rise of human life on Earth. The precise cause of this change is the imminent creation by technology of entities with greater than human intelligence. There are several means by which science may achieve this breakthrough (and this is another reason for having confidence that the event will occur):

- o There may be developed computers that are "awake" and superhumanly intelligent. (To date, there has been much controversy as to whether we can create human equivalence in a machine. But if the answer is "yes, we can", then there is little doubt that beings more intelligent can be constructed shortly thereafter.)

- o Large computer networks (and their associated users) may "wake up" as a superhumanly intelligent entity.

- o Computer/human interfaces may become so intimate that users may reasonably be considered superhumanly intelligent.

- o Biological science may provide means to improve natural human intellect.

The first three possibilities depend in large part on improvements in computer hardware. Progress in computer hardware has followed an amazingly steady curve in the last few decades¹. Based largely on this trend, I believe that the creation of greater than human intelligence will occur during the next thirty years. (Charles Platt² has pointed out that AI enthusiasts have been making claims like this for the last thirty years. Just so I'm not guilty of a relative-time ambiguity, let me more specific: I'll be surprised if this event occurs before 2005 or after 2030.)

What are the consequences of this event? When greater-than-human intelligence drives progress, that progress will be much more rapid. In fact, there seems no reason why progress itself would not involve the creation of still more intelligent entities — on a still-shorter time scale. The best analogy that I see is with the evolutionary past: Animals can adapt to

1. Moravec, Hans, *Mind Children*, Harvard University Press, 1988.

2. Platt, Charles, Private Communication.

problems and make inventions, but often no faster than natural selection can do its work — the world acts as its own simulator in the case of natural selection. We humans have the ability to internalize the world and conduct "what if's" in our heads; we can solve many problems thousands of times faster than natural selection. Now, by creating the means to execute those simulations at much higher speeds, we are entering a regime as radically different from our human past as we humans are from the lower animals.

From the human point of view this change will be a throwing away of all the previous rules, perhaps in the blink of an eye, an exponential runaway beyond any hope of control. Developments that before were thought might only happen in "a million years" (if ever) will likely happen in the next century. (In ³, Greg Bear paints a picture of the major changes happening in a matter of hours.)

I think it's fair to call this event a singularity ("the Singularity" for the purposes of this paper). It is a point where our old models must be discarded and a new reality rules. As we move closer to this point, it will loom vaster and vaster over human affairs till the notion becomes a commonplace. Yet when it finally happens it may still be a great surprise and a greater unknown. In the 1950s there were very few who saw it: Stan Ulam ⁴ paraphrased John von Neumann as saying:

One conversation centered on the ever accelerating progress of technology and changes in the mode of human life, which gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue.

Von Neumann even uses the term singularity, though it appears he is thinking of normal progress, not the creation of superhuman intellect. (For me, the superhumanity is the essence of the Singularity. Without that we would get a glut of technical riches, never properly absorbed (see ⁵).)

3.Bear, Greg, "Blood Music", Analog Science Fiction-Science Fact, June, 1983. Expanded into the novel Blood Music, Morrow, 1985.

4.Ulam, S., Tribute to John von Neumann, *Bulletin of the American Mathematical Society*, vol 64, nr 3, part 2, May 1958, pp1-49.

5.Stent, Gunther S., *The Coming of the Golden Age: A View of the End of Progress*, The Natural History Press, 1969.

In the 1960s there was recognition of some of the implications of superhuman intelligence. I. J. Good wrote ⁶ :

Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion," and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control... . It is more probable than not that, within the twentieth century, an ultraintelligent machine will be built and that it will be the last invention that man need make.

Good has captured the essence of the runaway, but does not pursue its most disturbing consequences. Any intelligent machine of the sort he describes would not be humankind's "tool" — any more than humans are the tools of rabbits or robins or chimpanzees.

Through the '60s and '70s and '80s, recognition of the cataclysm spread ^{7 8 9 10} . Perhaps it was the science-fiction writers who felt the first concrete impact. After all, the "hard" science-fiction writers are the ones who try to write specific stories about all that technology may do for us. More and more, these writers felt an opaque wall across the future. Once, they could put such fantasies millions of years in the future ¹¹ . Now they saw that their most diligent extrapolations resulted in the

6. Good, I. J., "Speculations Concerning the First Ultraintelligent Machine", in *Advances in Computers*, vol 6, Franz L. Alt and Morris Rubinoff, eds, pp31-88, 1965, Academic Press.

7. Vinge, Vernor, "Bookworm, Run!", *Analog*, March 1966, pp8-40. Reprinted in *True Names and Other Dangers*, Vernor Vinge, Baen Books, 1987.

8. Alfve'n, Hannes, writing as Olof Johanneson, *The End of Man?*, Award Books, 1969 earlier published as "The Tale of the Big Computer", Coward-McCann, translated from a book copyright 1966 Albert Bonniers Forlag AB with English translation copyright 1966 by Victor Gollanz, Ltd.

9. Vinge, Vernor, First Word, *Omni*, January 1983, p10.

10. Bear, Greg, "Blood Music", *Analog Science Fiction-Science Fact*, June, 1983. Expanded into the novel *Blood Music*, Morrow, 1985.

11. Stapledon, Olaf, *The Starmaker*, Berkley Books, 1961 (but from the date on forward, probably written before 1937).

unknowable ... soon. Once, galactic empires might have seemed a Post-Human domain. Now, sadly, even interplanetary ones are.

What about the '90s and the '00s and the '10s, as we slide toward the edge? How will the approach of the Singularity spread across the human world view? For a while yet, the general critics of machine sapience will have good press. After all, till we have hardware as powerful as a human brain it is probably foolish to think we'll be able to create human equivalent (or greater) intelligence. (There is the far-fetched possibility that we could make a human equivalent out of less powerful hardware, if we were willing to give up speed, if we were willing to settle for an artificial being who was literally slow ¹² . But it's much more likely that devising the software will be a tricky process, involving lots of false starts and experimentation. If so, then the arrival of self-aware machines will not happen till after the development of hardware that is substantially more powerful than humans' natural equipment.)

But as time passes, we should see more symptoms. The dilemma felt by science fiction writers will be perceived in other creative endeavors. (I have heard thoughtful comic book writers worry about how to have spectacular effects when everything visible can be produced by the technologically commonplace.) We will see automation replacing higher and higher level jobs. We have tools right now (symbolic math programs, cad/cam) that release us from most low-level drudgery. Or put another way: The work that is truly productive is the domain of a steadily smaller and more elite fraction of humanity. In the coming of the Singularity, we are seeing the predictions of true technological unemployment finally come true.

Another symptom of progress toward the Singularity: ideas themselves should spread ever faster, and even the most radical will quickly become commonplace. When I began writing science fiction in the middle '60s, it seemed very easy to find ideas that took decades to percolate into the cultural consciousness; now the lead time seems more like eighteen months. (Of course, this could just be me losing my

12.Vinge, Vernor, "True Names", *Binary Star Number 5*, Dell, 1981. Reprinted in *True Names and Other Dangers*, Vernor Vinge, Baen Books, 1987.

imagination as I get old, but I see the effect in others too.) Like the shock in a compressible flow, the Singularity moves closer as we accelerate through the critical speed.

And what of the arrival of the Singularity itself? What can be said of its actual appearance? Since it involves an intellectual runaway, it will probably occur faster than any technical revolution seen so far. The precipitating event will likely be unexpected — perhaps even to the researchers involved. ("But all our previous models were catatonic! We were just tweaking some parameters... .") If networking is widespread enough (into ubiquitous embedded systems), it may seem as if our artifacts as a whole had suddenly awakened.

And what happens a month or two (or a day or two) after that? I have only analogies to point to: The rise of humankind. We will be in the Post-Human era. And for all my rampant technological optimism, sometimes I think I'd be more comfortable if I were regarding these transcendental events from one thousand years remove ... instead of twenty.

Can the Singularity be Avoided?

Well, maybe it won't happen at all: Sometimes I try to imagine the symptoms that we should expect to see if the Singularity is not to develop. There are the widely respected arguments of Penrose ¹³ and Searle ¹⁴ against the practicality of machine sapience. In August of 1992, Thinking Machines Corporation held a workshop to investigate the question "How We Will Build a Machine that Thinks" ¹⁵ . As you might guess from the workshop's title, the participants were not especially supportive of the arguments against machine intelligence. In fact, there was general agreement that minds can exist on nonbiological substrates and that algorithms are of central importance to the existence of minds. However, there was much debate about the raw hardware power that is present in organic brains. A minority felt that the largest 1992 computers were within three orders of magnitude of the power of the human brain. The majority of the participants agreed with Moravec's estimate ¹⁶ that we are ten to forty years away from hardware parity. And yet there was another minority who pointed to ^{17 18} , and conjectured that the computational competence of single neurons may be far higher than generally believed. If so, our present computer hardware might be as much as ten orders of magnitude short of the equipment we carry around in our heads. If this is true (or for that matter, if the Penrose or Searle critique is valid), we might never see a Singularity. Instead, in the early '00s we would find our hardware

13. Penrose, Roger, *The Emperor's New Mind*, Oxford University Press, 1989.

14. Searle, John R., "Minds, Brains, and Programs", in *The Behavioral and Brain Sciences*, vol 3, Cambridge University Press, 1980. The essay is reprinted in *The Mind's I*, edited by Douglas R. Hofstadter and Daniel C. Dennett, Basic Books, 1981 (my source for this reference). This reprinting contains an excellent critique of the Searle essay.

15. Thearling, Kurt, "How We Will Build a Machine that Thinks", a workshop at Thinking Machines Corporation, August 24-26, 1992. Personal Communication.

16. Moravec, Hans, *Mind Children*, Harvard University Press, 1988.

17. Conrad, Michael *et al.*, "Towards an Artificial Brain", *BioSystems*, vol 23, pp175-218, 1989.

18. Rasmussen, S. *et al.*, "Computational Connectionism within Neurons: a Model of Cytoskeletal Automata Subserving Neural Networks", in *Emergent Computation*, Stephanie Forrest, ed., pp428-449, MIT Press, 1991.

performance curves beginning to level off — this because of our inability to automate the design work needed to support further hardware improvements. We'd end up with some very powerful hardware, but without the ability to push it further. Commercial digital signal processing might be awesome, giving an analog appearance even to digital operations, but nothing would ever "wake up" and there would never be the intellectual runaway which is the essence of the Singularity. It would likely be seen as a golden age ... and it would also be an end of progress. This is very like the future predicted by Gunther Stent. In fact, on page 137 of ¹⁹, Stent explicitly cites the development of transhuman intelligence as a sufficient condition to break his projections.

But if the technological Singularity can happen, it will. Even if all the governments of the world were to understand the "threat" and be in deadly fear of it, progress toward the goal would continue. In fiction, there have been stories of laws passed forbidding the construction of "a machine in the likeness of the human mind" ²⁰. In fact, the competitive advantage — economic, military, even artistic — of every advance in automation is so compelling that passing laws, or having customs, that forbid such things merely assures that someone else will get them first.

Eric Drexler ²¹ has provided spectacular insights about how far technical improvement may go. He agrees that superhuman intelligences will be available in the near future — and that such entities pose a threat to the human status quo. But Drexler argues that we can confine such transhuman devices so that their results can be examined and used safely. This is I. J. Good's ultraintelligent machine, with a dose of caution. I argue that confinement is intrinsically impractical. For the case of physical confinement: Imagine yourself locked in your home with only limited data access to the outside, to your masters. If those masters thought at a rate — say — one million times slower than you, there is little doubt that over a period of years

19. Stent, Gunther S., *The Coming of the Golden Age: A View of the End of Progress*, The Natural History Press, 1969.

20. Herbert, Frank, *Dune*, Berkley Books, 1985. However, this novel was serialized in *Analog Science Fiction-Science Fact* in the 1960s.

21. Drexler, K. Eric, *Engines of Creation*, Anchor Press/Doubleday, 1986.

(your time) you could come up with "helpful advice" that would incidentally set you free. (I call this "fast thinking" form of superintelligence "weak superhumanity". Such a "weakly superhuman" entity would probably burn out in a few weeks of outside time. "Strong superhumanity" would be more than cranking up the clock speed on a human-equivalent mind. It's hard to say precisely what "strong superhumanity" would be like, but the difference appears to be profound. Imagine running a dog mind at very high speed. Would a thousand years of doggy living add up to any human insight? (Now if the dog mind were cleverly rewired and then run at high speed, we might see something different... .) Many speculations about superintelligence seem to be based on the weakly superhuman model. I believe that our best guesses about the post-Singularity world can be obtained by thinking on the nature of strong superhumanity. I will return to this point later in the paper.)

Another approach to confinement is to build rules into the mind of the created superhuman entity (for example, Asimov's Laws ²²). I think that any rules strict enough to be effective would also produce a device whose ability was clearly inferior to the unfettered versions (and so human competition would favor the development of the those more dangerous models). Still, the Asimov dream is a wonderful one: Imagine a willing slave, who has 1000 times your capabilities in every way. Imagine a creature who could satisfy your every safe wish (whatever that means) and still have 99.9% of its time free for other activities. There would be a new universe we never really understood, but filled with benevolent gods (though one of my wishes might be to become one of them).

If the Singularity can not be prevented or confined, just how bad could the Post-Human era be? Well ... pretty bad. The physical extinction of the human race is one possibility. (Or as Eric Drexler put it of nanotechnology: Given all that such technology can do, perhaps governments would simply decide that they no longer need citizens!). Yet physical extinction may not be the scariest possibility. Again, analogies: Think of the different ways we relate to animals. Some of the crude physical

22. Asimov, Isaac, "Runaround", *Astounding Science Fiction*, March 1942, p94. Reprinted in *Robot Visions*, Isaac Asimov, ROC, 1990. Asimov describes the development of his robotics stories in this book.

abuses are implausible, yet... . In a Post-Human world there would still be plenty of niches where human equivalent automation would be desirable: embedded systems in autonomous devices, self-aware daemons in the lower functioning of larger sentients. (A strongly superhuman intelligence would likely be a Society of Mind ²³ with some very competent components.) Some of these human equivalents might be used for nothing more than digital signal processing. They would be more like whales than humans. Others might be very human-like, yet with a one-sidedness, a dedication that would put them in a mental hospital in our era. Though none of these creatures might be flesh-and-blood humans, they might be the closest things in the new environment to what we call human now. (I. J. Good had something to say about this, though at this late date the advice may be moot: Good ²⁴ proposed a "Meta-Golden Rule", which might be paraphrased as "Treat your inferiors as you would be treated by your superiors." It's a wonderful, paradoxical idea (and most of my friends don't believe it) since the game-theoretic payoff is so hard to articulate. Yet if we were able to follow it, in some sense that might say something about the plausibility of such kindness in this universe.)

I have argued above that we cannot prevent the Singularity, that its coming is an inevitable consequence of the humans' natural competitiveness and the possibilities inherent in technology. And yet ... we are the initiators. Even the largest avalanche is triggered by small things. We have the freedom to establish initial conditions, make things happen in ways that are less inimical than others. Of course (as with starting avalanches), it may not be clear what the right guiding nudge really is:

23. Minsky, Marvin, *Society of Mind*, Simon and Schuster, 1985.

24. Good, I. J., [Help! I can't find the source of Good's Meta-Golden Rule, though I have the clear recollection of hearing about it sometime in the 1960s. Through the help of the net, I have found pointers to a number of related items. G. Harry Stine and Andrew Haley have written about metalaw as it might relate to extraterrestrials: G. Harry Stine, "How to Get along with Extraterrestrials ... or Your Neighbor", *Analog Science Fact- Science Fiction*, February, 1980, p39-47.]

Other Paths to the Singularity: Intelligence Amplification

When people speak of creating superhumanly intelligent beings, they are usually imagining an AI project. But as I noted at the beginning of this paper, there are other paths to superhumanity. Computer networks and human-computer interfaces seem more mundane than AI, and yet they could lead to the Singularity. I call this contrasting approach Intelligence Amplification (IA). IA is something that is proceeding very naturally, in most cases not even recognized by its developers for what it is. But every time our ability to access information and to communicate it to others is improved, in some sense we have achieved an increase over natural intelligence. Even now, the team of a PhD human and good computer workstation (even an off-net workstation!) could probably max any written intelligence test in existence.

And it's very likely that IA is a much easier road to the achievement of superhumanity than pure AI. In humans, the hardest development problems have already been solved. Building up from within ourselves ought to be easier than figuring out first what we really are and then building machines that are all of that. And there is at least conjectural precedent for this approach. Cairns-Smith²⁵ has speculated that biological life may have begun as an adjunct to still more primitive life based on crystalline growth. Lynn Margulis (in²⁶ and elsewhere) has made strong arguments that mutualism is a great driving force in evolution.

Note that I am not proposing that AI research be ignored or less funded. What goes on with AI will often have applications in IA, and vice versa. I am suggesting that we recognize that in network and interface research there is something as profound (and potential wild) as Artificial Intelligence. With that insight, we may see projects that are not as directly applicable as conventional interface and network design work, but which serve to advance us toward the Singularity along the IA path.

25.Cairns-Smith, A. G., *Seven Clues to the Origin of Life*, Cambridge University Press, 1985.

26.Margulis, Lynn and Dorion Sagan, *Microcosmos, Four Billion Years of Evolution from Our Microbial Ancestors*, Summit Books, 1986.

Here are some possible projects that take on special significance, given the IA point of view:

- o Human/computer team automation: Take problems that are normally considered for purely machine solution (like hill-climbing problems), and design programs and interfaces that take a advantage of humans' intuition and available computer hardware. Considering all the bizarreness of higher dimensional hill-climbing problems (and the neat algorithms that have been devised for their solution), there could be some very interesting displays and control tools provided to the human team member.

- o Develop human/computer symbiosis in art: Combine the graphic generation capability of modern machines and the esthetic sensibility of humans. Of course, there has been an enormous amount of research in designing computer aids for artists, as labor saving tools. I'm suggesting that we explicitly aim for a greater merging of competence, that we explicitly recognize the cooperative approach that is possible. Karl Sims²⁷ has done wonderful work in this direction.

- o Allow human/computer teams at chess tournaments. We already have programs that can play better than almost all humans. But how much work has been done on how this power could be used by a human, to get something even better? If such teams were allowed in at least some chess tournaments, it could have the positive effect on IA research that allowing computers in tournaments had for the corresponding niche in AI.

- o Develop interfaces that allow computer and network access without requiring the human to be tied to one spot, sitting in front of a computer. (This is an aspect of IA that fits so well with known economic advantages that lots of effort is already being spent on it.)

- o Develop more symmetrical decision support systems. A popular research/product area in recent years has been decision support systems. This is a form of IA, but may be too focussed on systems that are oracular. As much as the program

27.Sims, Karl, "Interactive Evolution of Dynamical Systems", Thinking Machines Corporation, Technical Report Series (published in *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, Paris, MIT Press, December 1991.

giving the user information, there must be the idea of the user giving the program guidance.

- o Use local area nets to make human teams that really work (ie, are more effective than their component members). This is generally the area of "groupware", already a very popular commercial pursuit. The change in viewpoint here would be to regard the group activity as a combination organism. In one sense, this suggestion might be regarded as the goal of inventing a "Rules of Order" for such combination operations. For instance, group focus might be more easily maintained than in classical meetings. Expertise of individual human members could be isolated from ego issues such that the contribution of different members is focussed on the team project. And of course shared data bases could be used much more conveniently than in conventional committee operations. (Note that this suggestion is aimed at team operations rather than political meetings. In a political setting, the automation described above would simply enforce the power of the persons making the rules!)

- o Exploit the worldwide Internet as a combination human/machine tool. Of all the items on the list, progress in this is proceeding the fastest and may run us into the Singularity before anything else. The power and influence of even the present-day Internet is vastly underestimated. For instance, I think our contemporary computer systems would break under the weight of their own complexity if it weren't for the edge that the USENET "group mind" gives the system administration and support people! The very anarchy of the worldwide net development is evidence of its potential. As connectivity and bandwidth and archive size and computer speed all increase, we are seeing something like Lynn Margulis' ²⁸ vision of the biosphere as data processor recapitulated, but at a million times greater speed and with millions of humanly intelligent agents (ourselves).

The above examples illustrate research that can be done within the context of contemporary computer science departments. There are other paradigms. For example, much of the work in Artificial Intelligence and neural nets would benefit

28. Margulis, Lynn and Dorion Sagan, *Microcosmos, Four Billion Years of Evolution from Our Microbial Ancestors*, Summit Books, 1986.

from a closer connection with biological life. Instead of simply trying to model and understand biological life with computers, research could be directed toward the creation of composite systems that rely on biological life for guidance or for the providing features we don't understand well enough yet to implement in hardware. A long-time dream of science-fiction has been direct brain to computer interfaces^{29 30}. In fact, there is concrete work that can be done (and is being done) in this area:

- o Limb prosthetics is a topic of direct commercial applicability. Nerve to silicon transducers can be made³¹. This is an exciting, near-term step toward direct communication.

- o Direct links into brains seem feasible, if the bit rate is low: given human learning flexibility, the actual brain neuron targets might not have to be precisely selected. Even 100 bits per second would be of great use to stroke victims who would otherwise be confined to menu-driven interfaces.

- o Plugging in to the optic trunk has the potential for bandwidths of 1 Mbit/second or so. But for this, we need to know the fine-scale architecture of vision, and we need to place an enormous web of electrodes with exquisite precision. If we want our high bandwidth connection to be in addition to what paths are already present in the brain, the problem becomes vastly more intractable. Just sticking a grid of high-bandwidth receivers into a brain certainly won't do it. But suppose that the high-bandwidth grid were present while the brain structure was actually setting up, as the embryo develops. That suggests:

- o Animal embryo experiments. I wouldn't expect any IA success in the first years of such research, but giving developing brains access to complex simulated neural structures might be very interesting to the people who study how the embryonic brain develops. In the long run, such experiments might

29. Anderson, Poul, "Kings Who Die", *If*, March 1962, p8-36. Reprinted in *Seven Conquests*, Poul Anderson, MacMillan Co., 1969.

30. Vinge, Vernor, "Bookworm, Run!", *Analog*, March 1966, pp8-40. Reprinted in *True Names and Other Dangers*, Vernor Vinge, Baen Books, 1987.

31. Kovacs, G. T. A. *et al.*, "Regeneration Microelectrode Array for Peripheral Nerve Recording and Stimulation", *IEEE Transactions on Biomedical Engineering*, v 39, n 9, pp 893-902.

produce animals with additional sense paths and interesting intellectual abilities. Originally, I had hoped that this discussion of IA would yield some clearly safer approaches to the Singularity. (After all, IA allows our participation in a kind of transcendence.) Alas, looking back over these IA proposals, about all I am sure of is that they should be considered, that they may give us more options. But as for safety ... well, some of the suggestions are a little scary on their face. One of my informal reviewers pointed out that IA for individual humans creates a rather sinister elite. We humans have millions of years of evolutionary baggage that makes us regard competition in a deadly light. Much of that deadliness may not be necessary in today's world, one where losers take on the winners' tricks and are co-opted into the winners' enterprises. A creature that was built de novo might possibly be a much more benign entity than one with a kernel based on fang and talon. And even the egalitarian view of an Internet that wakes up along with all mankind can be viewed as a nightmare ³².

The problem is not simply that the Singularity represents the passing of humankind from center stage, but that it contradicts our most deeply held notions of being. I think a closer look at the notion of strong superhumanity can show why that is.

32. Swanwick Michael, *Vacuum Flowers*, serialized in *Isaac Asimov's Science Fiction Magazine*, December(?) 1986 - February 1987. Republished by Ace Books, 1988.

Strong Superhumanity and the Best We Can Ask for

Suppose we could tailor the Singularity. Suppose we could attain our most extravagant hopes. What then would we ask for: That humans themselves would become their own successors, that whatever injustice occurs would be tempered by our knowledge of our roots. For those who remained unaltered, the goal would be benign treatment (perhaps even giving the stay-behinds the appearance of being masters of godlike slaves). It could be a golden age that also involved progress (overleaping Stent's barrier). Immortality (or at least a lifetime as long as we can make the universe survive ^{33 34}) would be achievable.

But in this brightest and kindest world, the philosophical problems themselves become intimidating. A mind that stays at the same capacity cannot live forever; after a few thousand years it would look more like a repeating tape loop than a person. (The most chilling picture I have seen of this is in ³⁵ .) To live indefinitely long, the mind itself must grow ... and when it becomes great enough, and looks back ... what fellow-feeling can it have with the soul that it was originally? Certainly the later being would be everything the original was, but so much vastly more. And so even for the individual, the Cairns-Smith or Lynn Margulis notion of new life growing incrementally out of the old must still be valid.

This "problem" about immortality comes up in much more direct ways. The notion of ego and self-awareness has been the bedrock of the hardheaded rationalism of the last few centuries. Yet now the notion of self-awareness is under attack from the Artificial Intelligence people ("self-awareness and other delusions"). Intelligence Amplification undercuts our concept of ego from another direction. The post-Singularity world will involve extremely high-bandwidth networking. A central feature of strongly superhuman entities will likely be their ability to

33. Dyson, Freeman, "Physics and Biology in an Open Universe", *Review of Modern Physics*, vol 51, pp447-460, 1979.

34. Barrow, John D. and Frank J. Tipler, *The Anthropic Cosmological Principle*, Oxford University Press, 1986.

35. Niven, Larry, "The Ethics of Madness", *If*, April 1967, pp82-108. Reprinted in *Neutron Star*, Larry Niven, Ballantine Books, 1968.

communicate at variable bandwidths, including ones far higher than speech or written messages. What happens when pieces of ego can be copied and merged, when the size of a selfawareness can grow or shrink to fit the nature of the problems under consideration? These are essential features of strong superhumanity and the Singularity. Thinking about them, one begins to feel how essentially strange and different the Post-Human era will be — no matter how cleverly and benignly it is brought to be.

From one angle, the vision fits many of our happiest dreams: a time unending, where we can truly know one another and understand the deepest mysteries. From another angle, it's a lot like the worst- case scenario I imagined earlier in this paper.

Which is the valid viewpoint? In fact, I think the new era is simply too different to fit into the classical frame of good and evil. That frame is based on the idea of isolated, immutable minds connected by tenuous, low-bandwidth links. But the post-Singularity world does fit with the larger tradition of change and cooperation that started long ago (perhaps even before the rise of biological life). I think there are notions of ethics that would apply in such an era. Research into IA and high-bandwidth communications should improve this understanding. I see just the glimmerings of this now ³⁶. There is Good's Meta-Golden Rule; perhaps there are rules for distinguishing self from others on the basis of bandwidth of connection. And while mind and self will be vastly more labile than in the past, much of what we value (knowledge, memory, thought) need never be lost. I think Freeman Dyson has it right when he says ³⁷: "God is what mind becomes when it has passed beyond the scale of our comprehension."

[I wish to thank John Carroll of San Diego State University and Howard Davidson of Sun Microsystems for discussing the draft version of this paper with me.]

36.Vinge, Vernor, To Appear [:-)]

37.Dyson, Freeman, *Infinite in All Directions*, Harper && Row, 1988.



www.feedbooks.com
Food for the mind