

Technological Singularity by Vernor Vinge Magazine: Whole Earth Review
Issue: December 10, 1993
Title: Technological Singularity
Author: Vernor Vinge

Technological Singularity

by Vernor Vinge

Vernor Vinge's vision of a technological "singularity" in humanity's near future has haunted me since I first read of it in his science-fiction novel, *Marooned in Realtime* (1986). I'm persuaded that the acceleration of technology-acceleration is even now distorting human institutions and expectations, whether or not we are approaching a metaphorical "event horizon" beyond which everything becomes unrecognizable.

When I invited Vinge to write something about his current views on the singularity for the recent issue of *Whole Earth Review* that I guest-edited, he replied that he had just presented a paper on the subject for the VISION-21 Symposium, sponsored by the NASA Lewis Research Center and the Ohio Aerospace Institute. In due course he revised the piece and sent it along. I can think of no other technical paper that has so many references to science-fiction literature, as well it should.

Vinge is a mathematician at San Diego State University, specializing in distributed computing and computer architecture. One of his short stories, "True Names" (1981), is often mentioned along with John Brunner's *Shockwave Rider* and William Gibson's *Neuromancer* as an inspiration to the current generation of online computer pioneers. Vinge's two "Realtime" novels (combined in *Across Realtime* -- 1991) have been nominated for Hugo Awards, science fiction's top prize. His new novel, *A Fire Upon the Deep*, won the 1993 Hugo; it's reviewed on p. 95.

--Stewart Brand

TECHNOLOGICAL SINGULARITY

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

A slightly different 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. --Vernor Vinge

1. What Is The Singularity?

The acceleration of technological progress has been the central feature of this century. 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. Science may achieve this breakthrough by several means (and this is another reason for having confidence that the event will occur):

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

Large computer networks and their associated users may "wake up" as superhumanly intelligent entities.

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

Biological science may provide means to improve natural human intellect.

The first three possibilities depend on improvements in computer hardware. Progress in hardware has followed an amazingly steady curve in the last few decades. Based on this trend, I believe that the creation of greater-than-human intelligence will occur during the next thirty years. (Charles Platt has pointed out that AI enthusiasts have been making claims like this for thirty years. Just so I'm not guilty of a relative-time ambiguity, let me be 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 I see is to the evolutionary past: Animals can adapt to 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 could. 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.

This change will be a throwing-away of all the human rules, perhaps in the blink of an eye -- an exponential runaway beyond any hope of control. Developments that were thought might only happen in "a million years" (if ever) will likely happen in the next century.

It's fair to call this event a singularity ("the Singularity" for the purposes of this piece). It is a point where our old models must be discarded and a new reality rules, a point that will loom vaster and vaster over human affairs until the notion becomes a commonplace. Yet when it finally happens, it may still be a great surprise and a greater unknown. In the 1950s very few saw it: Stan Ulam 1 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.)

The 1960s saw recognition of some of the implications of superhuman intelligence. I. J. Good wrote:

Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of 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 he 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, robins, or chimpanzees.

Through the sixties and seventies and eighties, recognition of the cataclysm spread. 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. Now they saw that their most diligent extrapolations resulted in the unknowable . . . soon. Once, galactic empires might have seemed a Posthuman domain. Now, sadly, even interplanetary ones are.

What about the coming decades, 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, until 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 farfetched 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 that was literally slow. 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 until 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 comicbook writers worry about how to create 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. 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 will see 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.

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 by 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 Posthuman era. And for all my technological optimism, I think I'd be more comfortable if I were regarding these transcendental events from one thousand years' remove . . . instead of twenty.

2. Can the Singularity Be Avoided?

Well, maybe it won't happen at all: sometimes I try to imagine the symptoms we should expect to see if the Singularity is not to develop. There are the widely respected arguments of Penrose³ and Searle⁴ against the practicality of machine sapience. In August 1992, Thinking Machines Corporation held a workshop to investigate "How We Will Build a Machine That Thinks." 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 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 Hans Moravec's estimate⁵ that we are ten to forty years away from hardware parity. And yet there was an other minority who 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 performance curves beginning to level off -- 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 that 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,⁶ who 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. The competitive advantage -- economic, military, even artistic -- of every advance in automation is so compelling that forbidding such things merely assures that someone else will get them first.

Eric Drexler has provided spectacular insights about how far technical improvement may go.⁷ He agrees that superhuman intelligences will be available in the near future. But Drexler argues that we can confine such transhuman devices so that their results can be examined and used safely.

I argue that confinement is intrinsically impractical. 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 (your time) you could come up with a way to escape. 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? 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.

Another approach to confinement is to build rules into the mind of the created superhuman entity. I think that any rules strict enough to be effective would also produce a device whose ability was clearly inferior to the unfettered versions (so human competition would favor the development of the more dangerous models).

If the Singularity can not be prevented or confined, just how bad could the Posthuman 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. Think of the different ways we relate to animals. A Posthuman world would still have 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 Mind8 with some very competent components.) Some of these human equivalents might be used for nothing more than digital signal processing. Others might be very humanlike, yet with a onesidedness, 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 have argued above that we cannot prevent the Singularity, that its coming is an inevitable consequence of 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, to 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:

3. Other Paths to the Singularity

When people speak of creating superhumanly intelligent beings, they are usually imagining an AI project. But as I noted at the beginning of this article, there are other paths to superhumanity. Computer networks and human-computer interfaces seem more mundane than AI, yet they could lead to the Singularity. I call this contrasting approach Intelligence Amplification (IA). IA is proceeding very naturally, in most cases not even recognized for what it is by its developers. 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 Ph.D. 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 what we really are and then building machines that are all of that. And there is at least conjectural precedent for this approach. Cairns-Smith⁹ has speculated that biological life may have begun as an adjunct to still more primitive life based on crystalline growth. Lynn Margulis (in¹⁰ 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. AI advances 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 potentially 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.

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

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

Human/computer symbiosis in art: Combine the graphic generation capability of modern machines and the esthetic sensibility of humans. Of course, an enormous amount of research has gone into designing computer aids for artists. I'm suggesting that we explicitly aim for a greater merging of competence, that we explicitly recognize the cooperative approach that is possible. Karl Sims has done wonderful work in this direction.¹¹

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.

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

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 focused on systems that are oracular. As much as the program giving the user information, there must be the idea of the user giving the program guidance.

Local area nets to make human teams more effective than their component members. This is generally the area of "groupware"; the change

in viewpoint here would be to regard the group activity as a combination organism.

In one sense, this suggestion's goal might be to invent a "Rules of Order" for such combination operations. For instance, group focus might be more easily maintained than in classical meetings. Individual members' expertise could be isolated from ego issues so that the contribution of different members is focused on the team project. And of course shared databases could be used much more conveniently than in conventional committee operations.

The Internet as a combination human/machine tool. Of all the items on the list, progress in this is proceeding the fastest. The power and influence of the Internet are vastly underestimated. The very anarchy of the worldwide net's development is evidence of its potential. As connectivity, bandwidth, archive size, and computer speed all increase, we are seeing something like Lynn Margulis' 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 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 features we don't understand well enough yet to implement in hardware. A longtime dream of science fiction has been direct brain-to-computer interfaces. In fact, concrete work is being done in this area:

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

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.

Plugging into 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 add to the paths 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 as the brain structure was setting up, as the embryo developed. That suggests:

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, in the long run, produce animals with additional sense paths and interesting intellectual abilities.

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, about all I am sure of is that these proposals should be considered, that they may give us more options. But as

for safety -- some of the suggestions are a little scary on their face. 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 coopted into the winners' enterprises. A creature that was built de novo might possibly be a much more benign entity than one based on fang and talon.

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.

4. 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 occurred 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 (leaping Stent's barrier). Immortality (or at least a lifetime as long as we can make the universe survive) 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. 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? The later being would be everything the original was, but 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. 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 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 self-awareness 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 Posthuman 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 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 certain notions of ethics 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; 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, "God is what mind becomes when it has passed beyond the scale of our comprehension."12 |

1. Ulam, S., "Tribute to John von Neumann," Bulletin of the American Mathematical Society, vol. 64. no. 3, May 1958, pp. 1-49.

2. Good, I. J., "Speculations Concerning the First Ultraintelligent Machine," in Advances in Computers, vol 6, Franz L. Alt and Morris Rubinoff, eds., 31-88, 1965, Academic Press.

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

4. Searle, John R., "Minds, Brains, and Programs," in The Behavioral and Brain Sciences, vol. 3, Cambridge University Press, 1980.

5. Moravec, Hans, Mind Children, Harvard University Press, 1988.

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

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

8. Minsky, Marvin, Society of Mind, Simon and Schuster, 1985.

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

10. Margulis, Lynn and Dorian Sagan, Microcosmos: Four Billion Years of Evolution From Our Microbial Ancestors, Summit Books, 1986.

11. 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.

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

Other Sources

Alfvén, Hannes, writing as Olof Johanneson, The End of Man?, Award Books, 1969.

Anderson, Poul, "Kings Who Die," If, March 1962, 8-36.

Asimov, Isaac, "Runaround," Astounding Science Fiction, March 1942, 94.

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

Bear, Greg, "Blood Music," Analog Science Fiction-Science Fact, June, 1983.

Conrad, Michael, et al., "Towards an Artificial Brain," BioSystems, vol. 23, 175-218, 1989.

Dyson, Freeman, "Physics and Biology in an Open Universe," Review of Modern Physics, vol. 51, 447-460, 1979.

Herbert, Frank, Dune, Berkeley Books, 1985.

Kovacs, G. T. A., et al., "Regeneration Microelectrode Array for Peripheral Nerve Recording and Stimulation," IEEE Transactions on Biomedical Engineering, vol. 39, no. 9, 893-902.

Niven, Larry, "The Ethics of Madness," If, April 1967, 82-108.

Platt, Charles, private communication.

Rasmussen, S. et al., "Computational Connectionism within Neurons: a Model of Cytoskeletal Automata Subservicing Neural Networks," in Emergent Computation, Stephanie Forrest, ed., 428-449, MIT Press, 1991.

Stapledon, Olaf, The Starmaker, Berkeley Books, 1961.

Swanwick Michael, Vacuum Flowers, serialized in Isaac Asimov's Science Fiction Magazine, December 1986 - February 1987.

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

Vinge, Vernor, "Bookworm, Run!," Analog, March 1966, 8-40.

Vinge, Vernor, "True Names," Binary Star Number 5, Dell, 1981.

Vinge, Vernor, "First Word," Omni, January 1983, 10.

The contents of this file are copyright 1993 by the publisher
in whose directory this file appeared. Unauthorized copying
of this information is strictly forbidden. Please read the
general notice at the top menu of the Gopher Server for
the Electronic Newsstand. For information regarding reprints,
please send mail to REPRINTS@Enews.Com
