Select Font Size: A A A

Signs of the Singularity

By **Vernor Vinge**

This is part of *IEEE Spectrum*'s SPECIAL REPORT: THE SINGULARITY

IMAGE: BRYAN CHRISTIE DESIGN

I think it's likely that with technology we can in the fairly near future create or become creatures of more than human intelligence. Such a technological singularity would revolutionize our world, ushering in a posthuman epoch. If it were to happen a million years from now, no big deal. So what do I mean by "fairly near" future? In my 1993 essay, "The Coming Technological Singularity," I said I'd be surprised if the singularity had not happened by 2030. I'll stand by that claim, assuming we avoid the showstopping catastrophes—things like nuclear war, superplagues, climate crash—that we properly spend our anxiety upon.

In that event, I expect the singularity will come as some combination of the following:

The AI Scenario: We create superhuman artificial intelligence (AI) in computers.

The IA Scenario: We enhance human intelligence through human-to-computer interfaces—that is, we achieve intelligence amplification (IA).

The Biomedical Scenario: We directly increase our intelligence by improving the neurological operation of our brains.

The Internet Scenario: Humanity, its networks, computers, and databases become sufficiently effective to be considered a superhuman being.

The Digital Gaia Scenario: The network of embedded microprocessors becomes sufficiently effective to be considered a superhuman being.

The essays in this issue of *IEEE Spectrum* use similar definitions for the technological singularity but variously rate the notion from likely to totally bogus. I'm going to respond to arguments made in these essays and also mine them for signs of the oncoming singularity that we might track in the future.

Philosopher Alfred Nordmann criticizes the extrapolations used to argue for the singularity. Using trends for outright forecasting is asking for embarrassment. And yet there are a couple of

trends that at least raise the possibility of the technological singularity. The first is a very long-term trend, namely Life's tendency, across aeons, toward greater complexity. Some people see this as unstoppable progress toward betterment. Alas, one of the great insights of 20th-century natural science is that Nature can be the harshest of masters. What we call progress can fail. Still, in the absence of a truly terminal event (say, a nearby gamma-ray burst or another collision such as made the moon), the trend has muddled along in the direction we call forward. From the beginning, Life has had the ability to adapt for survival via natural selection of heritable traits. That computational scheme brought Life a long way, resulting in creatures that could reason about survival problems. With the advent of humankind, Life had a means of solving many problems much faster than natural selection.

In the last few thousand years, humans have begun the next step, creating tools to support cognitive function. For example, writing is an off-loading of memory function. We're building tools—computers, networks, database systems—that can speed up the processes of problem solving and adaptation. It's not surprising that some technology enthusiasts have started talking about possible consequences. Depending on our inventiveness—and our artifacts' inventiveness—there is the possibility of a transformation comparable to the rise of human intelligence in the biological world. Even if the singularity does not happen, we are going to have to put up with singularity enthusiasms for a long time.

Get used to it.

In recent decades, the enthusiasts have been encouraged by an enabling trend: the exponential improvement in computer hardware as described by Moore's Law, according to which the number of transistors per integrated circuit doubles about every two years. At its heart, Moore's Law is about inventions that exploit one extremely durable trick: optical lithography to precisely and rapidly emplace enormous numbers of small components. If the economic demand for improved hardware continues, it looks like Moore's Law can continue for some time—though eventually we'll need novel component technology (perhaps carbon nanotubes) and some new method of high-speed emplacement (perhaps self-assembly). But what about that economic demand? Here is the remarkable thing about Moore's Law: it enables improvement in communications, embedded logic, information storage, planning, and design—that is, in areas that are directly or indirectly important to almost all enterprise. As long as the software people can successfully exploit Moore's Law, the demand for this progress should continue.

> *The best answer to the question, "Will computers ever be as smart as humans?" is probably "Yes, but only briefly"*

Roboticist Hans Moravec may have been the first to draw a numerical connection between computer hardware trends and artificial intelligence. Writing in 1988, Moravec took his estimate of the raw computational power of the brain together with the rate of improvement in computer power and projected that by 2010 computer hardware would be available to support roughly human levels of performance. There are a number of reasonable objections to this line of argument. One objection is that Moravec may have radically underestimated the computational power of neurons. But even if his estimate is a few orders of magnitude too low, that will only delay the transition by a decade or two—assuming that Moore's Law holds.

Another roboticist, Rodney Brooks, suggests in this issue that computation may not even be the right metaphor for what the brain does. If we are profoundly off the mark about the nature of thought, then this objection could be a showstopper. But research that might lead to the singularity covers a much broader range than formal computation. There is great variety even in the pursuit of pure AI. In the next decade, those who credit Moravec's timeline begin to expect results. Interestingly powerful computers will become cheap enough for a thousand research groups to bloom. Some of these researchers will pursue the classic computational tradition that Brooks is doubting—and they may still carry the day. Others will be working on their own abstractions of natural mind functions—for instance, the theory that Christof Koch and Giulio Tononi discuss in their article. Some (very likely Moravec and Brooks himself) will be experimenting with robots that cope with many of the same issues that, for animals, eventually resulted in minds that plan and feel. Finally, there will be pure neurological researchers, modeling increasingly larger parts of biological brains *in silico*. Much of this research will benefit from improvements in our tools for imaging brain function and manipulating small regions of the brain.

But despite Moravec's estimate and all the ongoing research, we are far short of putting the hardware together successfully. In his essay, Brooks sets several intermediate challenges. Such goals can help us measure the progress that is being made. More generally, it would be good to have indicators and counterindicators to watch for. No single one would prove the case for

or against the singularity, but together they would be an ongoing guide for our assessment of the matter. Among the counterindicators (events arguing against the likelihood of the singularity) would be debacles of overweening software ambition: events ranging from the bankruptcy of a major retailer upon the failure of its new inventory management system to the defeat of network-centric war fighters by a transistor-free light infantry. A tradition of such debacles could establish limits on application complexity—independent of any claims about the power of the underlying hardware.

There are many possible positive indicators. The Turing Test—whether a human judge communicating by text alone can distinguish a computer posing as human from a real human—is a subtle but broad indicator. Koch and Tononi propose a version of the Turing Test for machine consciousness in which the computer is presented a scene and asked to "extract the gist of it" for evaluation by a human judge. One could imagine restricted versions of the Turing Test for other aspects of Mind, such as introspection and common sense.

As with past computer progress, the achievement of some goals will lead to interesting disputes and insights. Consider two of Brooks's challenges: manual dexterity at the level of a 6-year-old child and object-recognition capability at the level of a 2-year-old. Both tasks would be much easier if objects in the environment possessed sensors and effectors and could communicate. For example, the target of a robot's hand could provide location and orientation data, even URLs for specialized manipulation libraries. Where the target has effectors as well as sensors, it could cooperate in the solution of kinematics issues. By the standards of today, such a distributed solution would clearly be cheating. But embedded microprocessors are increasingly widespread. Their coordinated presence may become the assumed environment. In fact, such coordination is much like relationships that have evolved between living things.

There are more general indicators. Does the distinction between neurological and AI researchers continue to blur? Does cognitive biomimetics become a common source of performance improvement in computer applications? From an entirely different direction, consider economist Robin Hanson's "shoreline" metaphor for the boundary between those tasks that can be done by machines and those that can be done only by human beings. Once upon a time, there was a continent of human-only tasks. By the end of the 1900s, that continent had become an archipelago. We might recast much of our discussion in terms of the question, "Is any place on the archipelago safe from further inundation?" Perhaps we could track this process with an objective economic index—say, wages divided by world product. However much human wealth and welfare may increase, a sustained decline in the ratio of wages to world product would argue a decline in the human contribution to the economy.

Machine/network life-forms will be faster, more labile, and more varied than what we see in biology. Digital Gaia is a hint of how alien the possibilities are

Some indicators relate different areas of technological speculation. In his essay, physicist Richard A.L. Jones critiques molecular nanotechnology (MNT). Even moderate success with MNT could support Moore's Law long enough to absorb a number of order-of-magnitude errors in our estimates of the computing power of the brain. At the same time, some of the advanced applications that K. Eric Drexler describes—things like cell-repair machines—depend on awesome progress with software. Thus, while success with MNT probably does not need the technological singularity (or vice versa), each would be a powerful indicator for the other.

Several of the essays discuss the plausibility of mind uploads and consequent immortality for "our digitized psyches," ideas that have recently appeared in serious nonfiction, most notably Ray Kurzweil's *The Singularity Is Near*. As with nanotechnology, such developments aren't prerequisites for the singularity. On the other hand, the goal of enhancing human intelligence through human-computer interfaces (the IA Scenario) is both relevant and in view. Today a well-trained person with a suitably provisioned computer can look very smart indeed. Consider just a slightly more advanced setup, in which an Internet search capability plus math and modeling systems are integrated with a head-up display. The resulting overlays could give the user a kind of synthetic intuition about his or her surroundings. At a more intimate but still noninvasive level, DARPA's Cognitive Technology Threat Warning System is based on the idea of monitoring the user's mental activities and feeding the resulting analysis back to the user as a supplement to his or her own attention. And of course there are the researchers working with direct neural connections to machines. Larger numbers of implanted connections may allow selection for effective subsets of connections. The human and the machine sides can train to accommodate each other.

To date, research on neural prostheses has mainly involved hearing, vision, and communication.

Prostheses that could restore any cognitive function would be a very provocative indicator. In his essay, John Horgan discusses neural research, including that of T.W. Berger, into prostheses for memory function. In general, Horgan and I reach very different conclusions, but I don't think we have much disagreement about the facts; Horgan cites them to show how distant today's technology is from anything like the singularity—and I am saying, "Look here, these are the sorts of things we should track going forward, as signs of progress toward the singularity (or not)."

The Biomedical Scenario—directly improving the functioning of our own brains—has a lot of similarities to the IA Scenario, though computers would be only indirectly involved, in support of bioinformatics. In the near future, drugs for athletic ability may be only a small problem compared with drugs for intellect. If these mind drugs are not another miserable fad of uppers and downers, if they enable real improvements to memory and creativity, that would be a strong indicator for this scenario. Much further out—for both logistical and ethical reasons—is the possibility of embryo optimization and germ-line engineering. Biomedical enhancement, even the extreme varieties, probably does not scale very well; however, it might help biological minds maintain some influence over other progress.

Brooks suggests that the singularity might happen—and yet we might not notice. Of the scenarios I mentioned at the beginning of this essay, I think a pure Internet Scenario—where humanity plus its networks and databases become a superhuman being—is the most likely to leave room to argue about whether the singularity has happened or not. In this future, there might be all-but-magical scientific breakthroughs. The will of the people might manifest itself as a seamless transformation of demand and imagination into products and policy, with environmental and geopolitical disasters routinely finessed. And yet there might be no explicit evidence of a superhuman player.

A singularity arising from networks of embedded microprocessors—the Digital Gaia Scenario—would probably be less deniable, if only because of the palpable strangeness of the everyday world: reality itself would wake up. Though physical objects need not be individually sapient, most would know what they are, where they are, and be able to communicate with their neighbors (and so potentially with the world). Depending on the mood of the network, the average person might notice a level of convenience that simply looks like marvelously good luck. The Digital Gaia would be something beyond human intelligence, but nothing like human. In general, I suspect that machine/network life-forms will be faster, more labile, and more varied than what we see in biology. Digital Gaia is a hint of how alien the possibilities are.

In his essay, Hanson focuses on the economics of the singularity. As a result, he produces spectacular insights while avoiding much of the distracting weirdness. And yet weirdness necessarily leaks into the latter part of his discussion (even leaving Digital Gaia possibilities aside). AI at the human level would be a revolution in our worldview, but we can already create human-level intelligences; it takes between nine months and 21 years, depending on whom you're talking to. The consequences of creating human-level artificial intelligence would be profound, but it would still be explainable to present-day humans like you and me.

But what happens a year or two after that? The best answer to the question, "Will computers ever be as smart as humans?" is probably "Yes, but only briefly."

For most of us, the hard part is believing that machines could ever reach parity. If that does happen, then the development of superhuman performance seems very likely—and that is the singularity. In its simplest form, this might be achieved by "running the processor clock faster" on machines that were already at human parity. I call such creatures "weakly superhuman," since they should be understandable if we had enough time to analyze their behavior. Assuming Moore's Law muddles onward, minds will become steadily smarter. Would economics still be an important driver? Economics arises from limitations on resources. Personally, I think there will always be such limits, if only because Mind's reach will always exceed its grasp. However, what is scarce for the new minds and how they deal with that scarcity will be mostly opaque to us.

The period when economics could help us understand the new minds might last decades, perhaps corresponding to what Brooks describes as "a period, not an event." I'd characterize such a period as a soft takeoff into the singularity. Toward the end, the world would be seriously strange from the point of view of unenhanced humans.

A soft takeoff might be as gentle as changes that humanity has encountered in the past. But I think a hard takeoff is possible instead: perhaps the transition would be fast. One moment the world is like 2008, perhaps more heavily networked. People are still debating the possibility of the singularity. And then something...happens. I don't mean the accidental construction that Brooks describes. What I'm thinking of would probably be the result of intentional research, perhaps a group exploring the parameter space of their general theory. One of their experiments finally gets things right. The result transforms the world—in just a matter of hours. A hard takeoff into the singularity could resemble a physical explosion more than it does

technological progress.

---

*If the singularity happens, the world
passes beyond human ken*

I base the possibility of hard takeoff partly on the known potential of rapid malcode (remember the Slammer worm?) but also on an analogy: the most recent event of the magnitude of the technological singularity was the rise of humans within the animal kingdom. Early humans could effect change orders of magnitude faster than other animals could. If we succeed in building systems that are similarly advanced beyond us, we might experience a similar incredible runaway.

Whether the takeoff is hard or soft, the world beyond the singularity contains critters who surpass natural humans in just the ability that has so empowered us: intelligence. In human history, there have been a number of radical technological changes: the invention of fire, the development of agriculture, the Industrial Revolution. One might reasonably apply the term *singularity* to these changes. Each has profoundly transformed our world, with consequences that were largely unimagined beforehand. And yet those consequences could have been explained to earlier humans. But if the transformation discussed in this issue of *Spectrum* occurs, the world will become intrinsically unintelligible to the likes of us. (And that is why "singularity," as in "black hole singularity of physics," is *the* cool metaphor here.) If the singularity happens, we are no longer the apex of intellect. There will be superhumanly intelligent players, and much of the world will be to their design. Explaining that to one of us would be like trying to explain our world to a monkey.

Both Horgan and Nordmann express indignation that singularity speculation distracts from the many serious, real problems facing society. This is a reasonable position for anyone who considers the singularity to be bogus, but some form of the point should also be considered by less skeptical persons: if the singularity happens, the world passes beyond human ken. So isn't all our singularity chatter a waste of breath? There are reasons, some minor, some perhaps very important, for interest in the singularity. The topic has the same appeal as other great events in natural history (though I am more comfortable with such changes when they are at a paleontological remove). More practically, the notion of the singularity is simply a view of progress that we can use—along with other, competing, views—to interpret ongoing events and revise our local planning. And finally: if we are in a soft takeoff, then powerful components of superintelligence will be available well before any complete entity. Human planning and guidance could help avoid ghastliness, or even help create a world that is too good for us naturals to comprehend.

Horgan concludes that "the singularity is a religious rather than scientific vision." Brooks is more mellow, seeing "commonalities with religious beliefs" in many enthusiasts' ideas. I argue against Horgan's conclusion, but Brooks's observation is more difficult to dispute. If there were no other points to discuss, then those commonalities would be a powerful part of the skeptics' position. But there are other, more substantive arguments on both sides of the issue.

And of course, the spirituality card can be played against both skeptics and enthusiasts: Consciousness, intelligence, self-awareness, emotion—even their definitions have been debated since forever, by everyone from sophomores to great philosophers. Now, because of our computers, the applications that we are attempting, and the tools we have for observing the behavior of living brains, there is the possibility of making progress with these mysteries. Some of the hardest questions may be ill-posed, but we should see a continuing stream of partial answers and surprises. I expect that many successes will still be met by reasonable criticism of the form "Oh, but that's not really what intelligence is about" or "That method of solution is just an inflexible cheat." And yet for both skeptics and enthusiasts, this is a remarkable process. For the skeptic, it's a bit like subtractive sculpture, where step-by-step, each partial success is removing more dross, closing in on the ineffable features of Mind—a rather spiritual prospect! Of course, we may remove and remove and find that ultimately we are left with nothing but a pile of sand—and devices that are everything we are, and more. If that is the outcome, then we've got the singularity.

About the Author

VERNOR VINGE, who wraps up this issue, first used the term *singularity* to refer to the advent of superhuman intelligence while on a panel at the annual conference of the Association for the Advancement of Artificial Intelligence in 1982. Three of his books—*A Fire Upon the Deep* (1992), *A Deepness in the Sky* (1999), and *Rainbows End* (2006)—won the Hugo Award for best science-fiction novel of the year. From 1972 to 2000, Vinge taught math and computer science at San Diego State University.

To Probe Further

Vernor Vinge, "The Coming Technological Singularity," Vision-21 Symposium, NASA Lewis Research Center and the Ohio Aerospace Institute, 30 to 31 March 1993.

Vernor Vinge, "What If the Singularity Does NOT Happen," Seminars About Long-Term Thinking, the Long Now Foundation, 15 February 2007; see also the Long Now seminars, which are a wonderful resource for anyone who is interested in the long-term welfare of civilization.

Gordon E. Moore, "Cramming More Components Onto Integrated Circuits," *Electronics*, 19 April 1965. The accompanying illustration is an unsung virtue of the piece: clever, funny—and illustrating the limitations of even the most successful trend projections.

"International Technology Roadmap for Semiconductors," 2007 edition, Executive Summary.

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

DARPA solicitation, " Cognitive Technology Threat Warning System (CT2WS) Frequently Asked Questions," 8 April 2007.

Theodore W. Berger et al., "Brain-Implantable Biomimetic Electronics as a Neural Prosthesis for Hippocampal Memory Function," in *Toward Replacement Parts for the Brain*, Berger and Glanzman, eds., MIT Press, 2005.

David Moore et al., "The Spread of the Sapphire/Slammer Worm," 2003.

J. Storrs Hall, *Beyond AI: Creating the Conscience of the Machine*, Prometheus Books, 2007. Hall discusses strategies for reaching a favorable outcome in a soft takeoff.

For more articles, videos, and special features, go to The Singularity Special Report.