Observations

This Is Your Brain on Nanobots

Kevin Shapiro

In 1769, Wolfgang von Kempelen, a thirty-five-year-old Hungarian engineer, built the world's first chess computer—a chest-sized cabinet of gears and cogs behind which sat a wooden mannequin dressed as a Turk. After a sensational debut in 1770 at the imperial court in Vienna, where the clockwork Turk swiftly dispatched a courtier named Count Cobenzl, von Kempelen's contraption toured Europe for the next several decades. In celebrated matches, it defeated Benjamin Franklin, Catherine the Great, and Napoleon Bonaparte, among other luminaries of the age.

Of course, the Turk was a hoax; inside the cabinet, a human chessmaster controlled the turbaned dummy. Still, a mechanical man was not outside the realm of imagination in the Enlightenment era, and the mysterious automaton sparked serious debate over the promise and the limits of machines. In England, Robert Willis, who would later become a professor of applied mechanics at Cambridge, avowed his skepticism that a device like the Turk, with its paucity of movements and operations, could ever "usurp and exercise the faculties of the human mind"; chess, in his view, would remain "the province of the intellect alone." But there were also enthusiasts, like Willis's countryman Charles Babbage, who—after losing to the Turk in 1821—was inspired to design the first true mechanical computer. It was only a matter of time, Babbage reckoned, until a genuine chess-playing machine could be built.

Two centuries later, the terms of the debate over artificial intelligence (AI) remain essentially the same. Just as Babbage predicted, modern chess computers, like IBM's Deep Blue, are endowed with impressive powers of calculation. Nevertheless, skeptics of AI continue to contend that a computer's "intelligence" is not remotely humanlike and that a piece of hardware can succeed at chess only by brute force—which is to say, by generating and evaluating billions of possible board positions at every move.

For their part, AI enthusiasts point out that, only a few years ago, the skeptics doubted that even this was possible. Moreover, today's most powerful computers are less expensive and more flexible than Deep Blue, with improved software algorithms that recognize patterns rather than relying on overwhelming calculative might. For the enthusiasts, a machine capable of matching the computational ability of the human mind is just over the horizon.

Ray Kurzweil is perhaps today's foremost enthusiast. An inventor of reading machines, flat-bed scanners, music synthesizers, and speech-recognition software, among other bits of technological wizardry, Kurzweil is also known as an unabashed futurist and an animated proponent of artificial intelligence. In his first book, *The Age of Intelligent Machines* (1990), he predicted that a computer would defeat the world's chess champion by 1998. (In fact, Deep Blue defeated Garry Kasparov in 1997, a year ahead of schedule.) In a sequel, *The Age of Spiritual Machines* (1998), Kurzweil envisioned a not-too-distant age in which man and machine would become more and more alike. He prognosticated that computers

Kevin Shapiro *is a research fellow in neuroscience and a student at Harvard Medical School.*

would soon be wired into our bodies and brains, and that many of our interactions would take place in a domain of virtual reality, where it would be nearly impossible to distinguish human intelligence from its artificial counterpart.

Now, in *The Singularity Is Near*,* Kurzweil follows this line of thought to what he takes to be its logical end. As computers grow more and more powerful, he argues, artificial intelligence will not only surpass biological intelligence but will actually absorb it. Human civilization will be superseded by a hybrid civilization of men and computers; we will freely upload and download ideas, thoughts, and personalities, expanding our creative and scientific knowledge beyond currently imaginable limits. As time progresses, we will discard our organic habits and become beings of pure information, eventually saturating the entire universe with disembodied intelligence. Kurzweil calls this new paradigm the Singularity, and goes so far as to predict that it will be upon us within a few decades—around 2045, to be precise.

If this seems an impossibly brief time frame for such a radical transformation, Kurzweil responds that we have not absorbed the remarkably rapid pace of technological advance. In the first five chapters of *The Singularity Is Near*, he explains exactly how and when we will achieve the knowledge we need to attain the Singularity. In the last four chapters and epilogue, he explores the moral and philosophical ramifications of this sublime event, doing his best to calm the doubts of technophobes and more dismal prophets.

According to Kurzweil, what makes the Singularity inevitable is something he calls the Law of Accelerating Returns. Many of us are familiar with Moore's Law, the famous—and so far accurate—prediction by Gordon Moore, the chairman of Intel, that computer-chip power per unit of cost would double about every twelve months. In Kurzweil's view, this is merely one example of a much broader principle that applies to every mode of technology, from computer power and Internet bandwidth to DNA sequencing and brain scanning. All of these, he claims, are becoming exponentially more powerful (and cheaper) by the year.

Moreover, these trends will not be checked by the constraints of current technological know-how. The Law of Accelerating Returns guarantees that when refinements to a particular innovation, like the integrated circuit, begin to run up against physical and technical constraints, a new paradigm will emerge to continue the trend of exponential improvement. The integrated circuit itself replaced the transistor, which replaced the vacuum tube as the basic unit of computing; it will be replaced in turn by something else—a three-dimensional circuit, perhaps, or a DNA-based computer, or even a quantum device that performs calculations based on the properties of subatomic particles.

Practically speaking, the Law of Accelerating Returns means that we will soon have hardware with computational capacity that equals or exceeds that of the human brain. In fact, Kurzweil estimates that this will be available to consumers for about $1,000 by the year 2020, and he is confident we will not stop there. By about 2045—the projected date of the Singularity—$1,000 worth of computing power will generate about a billion times more intelligence per year than every human being alive today. We will have long since advanced beyond the prosaic laptop; computers will be integrated into our furniture, our clothing, our bodies, and our brains.

But hardware alone will not be enough to drive the Singularity. Even today's extraordinarily powerful machines have a difficult time performing some simple tasks, like recognizing an object wherever it happens to be located. This is not because they lack the computational resources, but because the software for computer vision, which derives largely from our understanding of how the brain sees, is still crude and immature.

Not to worry, says Kurzweil—the project of reverse-engineering the human brain is well under way, and should be complete by around 2029. Already, neuroscientists have constructed detailed models of brain regions as diverse as the auditory system, the cerebellum (which is responsible for fine motor coordination), and the hippocampus (which is thought to be important for memory). With improvements in the resolution of brain-imaging technologies, we will, Kurzweil imagines, be able to spy on the activity of individual brain cells without disturbing them. From there it is only a short step to reconstructing their patterns of information processing. Once that can be accomplished, we will have the tools necessary to create software that thinks just as the human brain does, if not better.

On another front, we are beginning to experiment with the integration of computer chips into neural networks. In fact, some of these ventures have already yielded fruit: cochlear implants send electrical impulses representing sound waves directly into the brains of the deaf, effectively replacing a damaged auditory nerve. Along similar lines, writes Kurzweil, we can envision retinal implants that will beam images into the mind's eye, and eventually nanobots—tiny, cell-sized computers—that will sit within the blood vessels of our brains, augmenting our thoughts and broadcasting them to a global Internet. Our nanobot-enhanced brains will be integrated seamlessly into a network of intelligence that is more artificial than biological, and human

* Viking, 672 pp., $29.95.

nature as we know it today will persist mostly as a historical legacy.

These, then, are the ingredients of the Singularity. First, computers with hardware vastly more efficient than the sloppy, error-ridden, decay-prone bundles of goo inside our skulls. Second, software that emulates the workings of the human mind, and in due course improves upon them. When these two ingredients are combined, the stage will be set for a radical and explosive reconfiguration of human life.

Kurzweil writes that the road to the Singularity—short though it may be—will run through three overlapping revolutions. We are already in the early stages of the first revolution, that of genetics and biotechnology, which will enable us to conquer disease and reprogram our bodies in anticipation of upcoming immortality.

But this is only the beginning. As Kurzweil sees it, biomedical tinkering will give way to nanotechnology. Just as nanobots will inhabit our brains, so too will they replace our blood cells, digestive systems, and virtually every other organ in our bodies. Self-replicating nanobot factories will also herald an end to poverty and the scarcity of resources, since they will be able to fabricate any desired object almost literally out of thin air. Once nanotechnology is mature enough, we will even replace our permanent bodies with custom-made avatars, changing our physical incarnation just as we might change clothes to suit our mood and the occasion.

The final step to the Singularity will be a revolution in robotics, which is to say, the creation of artificial intelligence that can outperform human intelligence. It becomes difficult even to envision the capabilities of this kind of "strong" AI, but Kurzweil writes that once it is achieved, there will be no limit to the expansion of a man-machine civilization. We will begin to utilize the information-processing abilities inherent in ordinary matter, from rocks to galaxies, until the universe itself starts to resemble an intelligent being.

The Singularity Is Near is in many ways a curious book. Formally, it is something of an awkward read: the flow of the text is encumbered by numerous sidebars, superfluous quotations, and bizarre dialogues between Kurzweil and several imagined interlocutors. The chapters themselves are filled with disjointed descriptions of recent scientific experiments and promising new technologies, which sometimes fail to distinguish the established from the speculative. More than anything else, however, one is nonplussed by Kurzweil's absolute confidence in his own predictions. Seemingly the only things that instill a modicum of uncertainty in him are those precluded by modern, pre-Singularity physics, like the possibility of altering the speed of light.

As for Kurzweil's vision of the future, in a sense it is nothing new. His so-called Singularity recalls the mid-20th-century transhumanism of the British biologist Julian Huxley, the Omega Point of the French Jesuit Pierre Teilhard de Chardin, and the noösphere of the Russian geochemist Vladimir Vernadsky. Although Kurzweil does not bother to acknowledge these philosophical predecessors, each of them predicted that human intelligence would one day transcend the limits of biology—and each did so without benefit of having witnessed the wonders of the microprocessor and the Internet.

Where *The Singularity Is Near* differs from these earlier speculations is in its detailed forecasts of the trajectory of technological development. But before taking these forecasts as certain and beginning to design our future virtual personas, it may be worth scrutinizing the basis for Kurzweil's remarkable predictions.

The Singularity Is Near is filled with graph after graph plotting developments in technology against a logarithmic ordinate, practically bludgeoning the reader with "proof" of the Law of Accelerating Returns. It is true that Kurzweil does attempt a more formal proof in an appendix (starting from some dubious axioms about "world knowledge" and "velocity of computation"), but the fact of the matter is that the Law of Accelerating Returns is not so much a law as an observation, and there is no particular reason to believe that it can be universally applied. On closer inspection, moreover, it becomes obvious that not all of the evidence adduced by Kurzweil on its behalf has been created equal.

It is one thing, for example, to demonstrate that computer processing has increased logarithmically in speed and accuracy over the past 30 years. It is quite another to draw a timeline positing a clockwork-like correspondence among such significant but disparate events as the origin of life (ca. 4 billion years ago), the divergence of humans and chimpanzees (ca. 10 million years ago), and the invention of the printing press (ca. 1450).

Some of Kurzweil's graphs, while not quite so preposterous as that, are uninterpretable. The ones plotting putative improvements in brain-imaging time and resolution are a case in point. There are numerous techniques for brain imaging in current use, each of which has its own advantages and disadvantages. Many branches of neuroscience, like the study of language and other higher cognitive functions, are not nearly mature enough to take advantage of the highest-resolution techniques. In any case, however, the advances in brain scanning have not yet translated into vastly increased knowledge about the human brain.

This is an important point, since reverse-engineering of the brain is in many ways the linchpin of Kurzweil's predictions. His claim that this project is well under way is based on a view of neuroscience that is astonishingly credulous—as if Kurzweil were taking at face value the boldest speculations made in press releases or in the Science section of the *New York Times*. Just because a particular neuroscientist claims to have explained an age-old enigma, like the origin of the emotions or the subtle inexactness of memory, does not mean that everyone else working in the field can fold up his tent and go home.

Is Kurzweil's neuro-boosterism a product more of educated over-enthusiasm than of naïveté? Unfortunately, his story is riddled with several obvious errors, conceptual and otherwise, that cast doubt on his scholarship. He claims, for instance, that pioneering work by David Hubel and Thorsten Wiesel in the 1960's showed that the brain could reorganize itself following damage or injury, as from a stroke. In fact, Hubel and Wiesel's work had nothing to do with brain damage. The two researchers are famous for having described how the visual system in kittens reorganized itself when deprived of *input*, as when the animals' eyes were sewn shut or forced surgically to squint. In fact, what they showed was that the brain is surprisingly *limited* in its ability to reorganize itself. Where the cats were concerned, once their initial wiring was laid down abnormally, it remained abnormal for the rest of their lives.

Contrary to Kurzweil's claims, we also have only the most elementary grasp of how the brain decodes sounds, stores memories, uses language, and coordinates movements. We are not even close to understanding disorders like Alzheimer's and epilepsy—indeed, we barely even know what causes them. And today's electronic neural "implants" are crude affairs: the devices that palliate Parkinson's disease simply deliver raw bursts of electrical activity to quiet the cells that generate maddening tremors and jerks. This is somewhat analogous to smacking a television set in order to improve its reception.

Nor are our extant simulations of human intelligence nearly so sophisticated as Kurzweil would have us believe. Although it is true, for example, that computers can now be programmed to read an electrocardiogram (EKG), one would be hard-pressed to find a physician who would trust his life—or anyone else's—to an automated machine. Computers can also compose music, but, aside from computer scientists, not many humans enjoy listening to it.

The tasks at which "weak" AI succeeds best are those that humans solve largely by trial and error, like finding the best way to allocate resources, or optimizing the arrangement of parts in a jet engine. But computers solve these problems in the same way humans do, although their trial-and-error cycles are much briefer. This is "intelligence" of a sort—but not the sort that most engages the interests of cognitive scientists, who are more concerned with understanding the basic mechanisms of seeing, speaking, recognizing, and remembering. When it comes to modeling the latter sort of processes, most neural networks are only a step ahead of von Kempelen's Turk: typically, what emerges is little more than a theorist's intuition of how the brain works, filtered through several layers of complicated programming.

Kurzweil's reputation as a prognosticator rests largely on successful predictions he has made in the past. In addition to anticipating the success of Deep Blue, he is also credited with foretelling the widespread dissemination of the Internet. Prescient though these predictions may have been, they were based on observations of technology that existed at the time they were made. Chess computers were good enough to defeat reasonably skilled humans by the time Deep Thought crushed the Scottish chessmaster David Levy in 1989, and the foundations for the Internet were laid in the ARPANET program plan of 1968. To bet against Ray Kurzweil in 1990, one would have had to be an ostrich.

By contrast, the technologies that are supposed to bring about the Singularity are today either still in their infancy (like neuroscience) or confined to the realm of theory and speculation (like nanotechnology). Is there a chance they will come to fruition? Of course—though probably not as completely or as quickly as Kurzweil imagines. However large and fast our computers, we are still a long, long way from the theoretical insights that will be needed to decode the blueprints of the mind. Cornucopian nanobots are likely to be even farther off. In the real world, we are still waiting for laptop batteries that last more than a few hours and practical alternatives to fossil fuels, both of which have been promised for quite a while.

Like Kurzweil himself, who reports that he takes 250 dietary supplements a day and undergoes twelve intravenous treatments a week, some of us are also waiting for the pharmaceutical panacea that will allow us to eat whatever we like and stop old age in its tracks. But even if the biological processes at play here were well-understood, it would still take decades before effective drugs came on the market. This is only in part because of the sluggishness of the FDA approval process, much lamented by Kurzweil. More decisive is the fact that drug development is an imperfect process, and even custom-tailored chemical cures have unpredictable side effects. Consider Vioxx, which was touted as a selective inhibitor of painful inflammation but turned out also to inhibit our natural defenses against heart attacks and stroke.

Perhaps, one day, some forward thinker like Kurzweil will be among the first to replace failing organs with permanent artificial substitutes, that is, ones that our bodies will not destroy or cover with useless fibrous tissue. For now, however, the only scientifically proven way to stay healthy is still the least technologically intensive. It consists of diet, exercise, and good sense.

Finally, there is the small matter of human nature. Kurzweil takes it for granted that the promise of increased longevity, wealth, and physical and intellectual capability will overwhelm any opposition to the total integration of man and machine. But even if it becomes technologically feasible, will we *want* to interact in virtual reality, except perhaps for purposes of entertainment? Will we wish to be permanently hooked in to a massive electronic consciousness? Psychology suggests otherwise, for reasons that have more to do with instinct than with intellect.

The virtual world posited by Kurzweil recalls the hypothetical experience machine imagined by the late philosopher Robert Nozick. This would be programmed by "super-duper neuropsychologists" to provide any sensation or experience whatsoever. As Nozick pointed out, however, if such a machine existed, most of us would not be inclined to plug into it. We are driven by our nature to interact in the real world, with our real bodies; an artificial reality—no matter how convincing—is not quite enough.

Nozick went on to argue that even if we could upgrade the experience machine into a result machine—which would transform us into whatever we desired to be and bring into being the appropriate causal results—we still would not want to plug in to it. However sophisticated the machine, it could not provide a satisfying alternative to unenhanced, unadulterated, individual human life. "What we desire," Nozick wrote, "is to live (an active verb) ourselves, in contact with reality. (And this, machines cannot do *for* us.)"

H.L. Mencken once quipped that physicists, who deal "habitually with objects and quantities far beyond the reach of the senses," have a peculiar tendency to

> fall into the ways of thinking of men dealing with objects and quantities that do not exist at all, e.g., theologians and metaphysicians. Thus, their speculations tend almost inevitably to depart from the field of true science, which is that of precise observation, and to become mere soaring in the empyrean.

One might make the same observation today about computer scientists, who also deal with the intangible as a matter of profession. It is perhaps no coincidence that futurists, whether of the enthusiastic or the skeptical stripe, are by and large computer scientists. They include not only Kurzweil but his frequent debating partner Bill Joy, founder of Sun Microsystems, who in recent years has gained a reputation as the Cassandra of futurist visionaries. Whatever their inclination, the essential question in their minds is not whether we will soon have nanobots and intelligent machines but how we will deal with them—whether we will be overrun by our creations, as Joy predicts, or transformed by them, as Kurzweil holds.

By contrast, neuroscientists, whose job it is to grapple with the mysteries of thought and behavior, tend to be rather more circumspect. Charles Sherrington, who pioneered modern neurophysiology early in the last century, was sanguine in his appraisal of the gulf between our understanding of biology and our understanding of psychology. As he remarked in *Man on His Nature* (1940):

> The mind is a something with such manifold variety, such fleeting changes, such countless nuances, such wealth of combinations, such heights and depths of mood, such sweeps of passion, such vistas of imagination, that the bald submission of some electrical potentials recognizable in nerve-centers as correlative to all these may seem to the special student of mind almost derisory. It is, further, more than mere lack of corresponding complexity which frustrates the comparison.

Despite the frankly extraordinary advances in brain science since Sherrington's time, this assessment remains as true today as it ever was. One can believe, in principle, that the operations of the brain are governed by physical laws, and yet still appreciate the profound depths of our ignorance about the causal links. True, research in neuroscience is growing exponentially, at least as judged by the proliferation of papers and scholarly journals. (Scientists, no less than real-estate speculators, are attracted to underdeveloped and promising areas.) And yet the Law of Accelerating Returns does not seem to apply to our basic understanding of the fundamental mysteries of cognition. If the transformation of human society depends on solving the puzzle of the human mind via reverse-engineering, then the Singularity is much farther away than Ray Kurzweil thinks.