GREY MATTERS

The Sum of All Thoughts: Prospects of Uploading the Mind to a Computer

WILLIAM P. CHESHIRE, JR., MD

I think the brain is like a program in the mind, which is like a computer. So it's theoretically possible to copy the brain on to a computer and so provide a form of life after death. – Stephen Hawking[1]

Abstract

Beginning with the premise that the human mind is fundamentally a computer, and extrapolating from the history of computer technology, which has yielded ever-increasing processing speeds, some futurists forecast a time when it may become possible to upload the human brain to a computer and thereby attain enhanced powers and a sort of immortality. Such predictions add new meaning to the idiom of having one's mind in a cloud. They also raise profound ethical questions. The suggestion that brain uploading could be achieved safely suggests unbridled hubris. The belief that human identity could be faithfully replicated in a machine is possible only within a reductionistic, hence inadequate, understanding of the human person. A hypothetical post-neuron future in silicon could never be more than a collection of inauthentic human representations.

Introduction

The last several decades have seen an explosion of information technology. The formalization of data in the language of computer programming combined with exponential increases in microelectronic processing speed have yielded computational machines that rival some of the cognitive capacities of the human brain. Regardless of whether the field of artificial intelligence succeeds in building machines that mimic or even surpass human cognition, its hypothetical basis already has implications for how people think about human intelligence.

One implication is that many people today view the brain as being essentially a computer. The analogy is increasingly evident in common language. The verb "to process," for example, which denotes a series of mechanical operations, is sometimes used to refer to reasoning or gaining an understanding of something. Whereas computers run on software, some say the brain thinks with "wetware."[2] After all, quipped a fictional neurosurgeon on the television drama *Three Pounds*, the brain is just "wires in a box."[3]

Assuming for the moment that brains and computers are functionally equivalent information processors that happen to utilize different hardware, then at every level the design of the brain would be a useful model for building better computers. In principle, it might be possible to translate neural signals into electronic currents and, by joining dendrites with nanofibers, to connect neural networks with silicon chips. No longer would there be a meaningful distinction between neuroscience and

computer engineering. Brain and machine would coalesce in their material unity. Accordingly, the futurist Ray Kurzweil predicts a day when computers will exceed human intelligence. He writes, "There are no inherent barriers to our being able to reverse engineer the operating principles of human intelligence that will become available in the decades ahead. . . . Once a computer achieves a human level of intelligence, it will necessarily soar past it."[4]

As computer technology has already profoundly shaped life in the 21st century, its future promises to be a source of both exhilaration and apprehension. Its grand achievements will ever be a fascinating subject for technical expertise as well as ethical reflection.

Assistant or Replacement

A further implication of the premise that brains and computers are identical is that, given a sufficiently robust computer chip and biomechanical interfacing, a computer might substitute for the brain. If the only meaningful difference between the two were seen as a choice of hardware, then a failing or aging brain might seek a more durable home within the circuits of the latest computer. Personal identity, memories, likes and dislikes, loves and fears, beliefs and aspirations—consciousness itself—would be reframed in a substrate of silicon, copper, plastic, and glass, enclosed perhaps in a polished aluminum pseudocranium.

The hypothetical procedure of transferring the mind to a computer is known as "uploading," which in computer engineering denotes the transfer of data from one computer system to a higher level computer. Uploading a human brain to a computer, writes Kurzweil, would mean "scanning all of its salient details and then reinstantiating those details into a suitably powerful computational substrate. This process would capture a person's entire personality, memory, skills, and history."[5] The aspect of such technology that he considers "the most compelling" would involve "the gradual but inexorable progression of humans themselves from biological to nonbiological."[6]

Joining the chorus of would-be robots is Nick Bostrom, who asserts that "Substrate is morally irrelevant, assuming it doesn't affect functionality or consciousness. It doesn't matter, from a moral point of view, whether somebody runs on silicon or biological neurons (just as it doesn't matter whether you have dark or pale skin). On the same grounds, that we reject racism and speciesiem, we should also reject carbon-chauvinism, or bioism."[7]

The Failed Comic Upload

Novel experiments seldom go precisely as planned, as I learned during my own first experiment with uploading, which was an experience that frames my evaluation of proposals to transfer a human mind into the silicon substrate of a computer circuitboard. I was seven years old—too young to know much about the exciting subjects that laid years ahead, like mathematics, science, medicine, and ethics, and long before the arrival of personal computers, cable television, or the Internet, but old enough to enjoy cartoons. That summer while visiting my grandmother in South Hill, Virginia, I noticed that some of the animated children's programs on her television

station were more interesting than the familiar ones I was used to watching back home in Charleston, South Carolina. So I devised a plan to persuade our local television station to carry them.

My writing skills at the time were rudimentary. I was incapable of advancing a cogent argument, had no contacts in the cartoon industry and no idea where cartoon animators practiced their craft, but I could draw. On the largest sheet of paper I could find, with meticulous ballpoint pen strokes and delicate crayon shading, I sketched each of the cartoon characters that appeared in the afternoon programs on the Virginia station, complete with block sequences telling captioned stories. My grandmother helped me to spell the names of the characters correctly. As television in those days was broadcast in black and white, I took creative license in adding color to the scenes. Once my mother and I returned to Charleston, at my insistence she took me to the WCSC television station on East Bay Street so that I could complete my mission, which was to convince the executives at the station to put those cartoon characters on the Charleston station so that my friends and I could tune in and watch them on a regular basis. With my pictures in hand, surely they would be able to find the corresponding programs, wherever one gets cartoon programs, and upload them to their broadcasts.

As the son of a newspaper editorial writer, and as a boy who regularly watched the evening news with my father, the media seemed as available to me as our own back yard. From my perspective, my father was head of it all, and whoever was in charge at the television station would certainly know that and agree to meet with me. We arrived at the television station without an appointment, and my mother pulled open the great glass doors as I ambled in, my little hands carefully holding the Crayola portfolio. For some reason the receptionist seemed puzzled, but after a short wait we were escorted back, and I was given a private audience with the host of the station's daily community affairs program. I entrusted him with my drawings, and he assured me that he would put them on television. I was elated. Getting things done in the real world, so it seemed, was easier than I had expected.

That afternoon when his daily program aired, my mother and I watched at home as he held my drawing before the camera, which zoomed in as he praised the youthful artwork. In less than a minute it was all over. That was it. I was dumbfounded. The problem was, by "putting my cartoons on television," we meant very different things.

The television host was extremely generous to air my sketch, but he misunderstood what I was unable to express in words, which was that I wanted the station to show the same programs that I enjoyed watching in Virginia. My drawing was merely a snapshot representation and not the actual animated cartoons that I hoped to see broadcast. It contained bits of information but was not the real thing. My drawings were only representations of animated cartoons, just as cartoons are only representations of living persons. To confuse one for the other leads to outcomes that at best are disappointing and at worst absurd.

To Upload or Not to Upload

The prospect of uploading a mind to a computer raises profound questions. Among them, at the level of engineering, is feasibility. The structural and functional complexities of the human brain pose an enormous challenge to proposals to transfer

its information content and internal networking to a machine. Each of the brain's hundred billion neurons connects with thousands of other neurons.[8] Kurzweil estimates that the brain comprises 10^{14} neural connections, which translates to some 10^{16} synaptic transactions per second.[9] And that estimate is just for neurons; it does not include the role of glial cells in shaping cognitive processes.[10]

Between neuroscience and computer science lies a huge chasm of incomplete knowledge. Stanley L. Jaki asserts that "No pronouncement on the identity of brain and computer should be paid serious attention when it is evident from the context that the gravity and extent of unsolved questions in brain research are systematically underplayed or simply overlooked."[11] Among the unresolved questions is that "The nervous system appears to be using a radically different system of notation from the ones we are familiar with in ordinary arithmetic and mathematics," for which reason "the mechanisms of the brain underlying the handling of information must be in principle unobservable to mechanistic investigation,"[11] if not also untranslatable into computer code.

Kurzweil attempts to resolve the complexity conundrum by appealing to Moore's Law, according to which historically computational speed has doubled approximately every two years. For Kurzweil, all practical objections based on current engineering limitations seem to vanish once Moore's Law is invoked. Extrapolated onto a historical graph that has risen exponentially, at first glance anything seems possible.

There is, however, more than Moore's law to consider. The claim that an exponential increase in computational power will continue in an unending trajectory, ultimately to surpass human intelligence, presupposes that there are no upper limits to the speed at which information can be transmitted. In reality, however, the laws of physics do impose physical constraints on signaling. With increasing miniaturization, physical and chemical interactions behave differently, because it is the quirky laws of quantum mechanics that govern the interactions of matter at the nanoscale.

The appeal to Moore's law as an engine that inevitably will merge human with computer intelligence also presupposes that human thought is fully reducible to mechanical processes. The assertion of reductionism, however, is not a scientific claim but a metaphysical one. The methodology of science, which considers only what can be known through empirical investigation of material phenomena, cannot prove through empirical investigation that there is nothing more to the human mind than science can measure in the brain.[12] Science, therefore, lacks the philosophical basis for assuring those who would consider uploading their brains that the process would retain their true selves.

Aside from such limits, another feature that is often overlooked is that dendritic connections between neurons are extraordinarily tiny and fragile. No current or foreseeable technology could capture an individual's entire neuronal architecture, let alone information stored deep inside neurons within the molecular configuration of nucleic acids, without destroying every detail of the brain, if not also the desired information itself, in the extraction process. Kurzweil's prediction that someone will invent nanobots that will somehow solve that problem[13] finds no plausible support in engineering theory. These would be the nanobots not of science but of comic strips.

Another interesting problem of mind uploading is that it challenges concepts of personal identity. Copying one's brain onto a silicon substrate would seem to create

an additional person having the same identity. The paradox would not be resolved by technology that in its course destroys the original brain, because once one's complete identity were to exist within a computer, it could be copied and recopied into a limitless number of other computers. Ray Kurzweil imagines such a scenario: "You could even scan and copy me while I was sleeping. If you come to me in the morning and say, 'Good news, Ray, we've successfully reinstantiated you into a more durable substrate, so we won't be needing your old body and brain anymore,' I may beg to differ."[14]

In a further examination of this paradox of multiple exact replicas laying claim to the same identity, Donald MacKay argues that it would seem "absurd to suggest that what identifies you is simply the information-flow pattern in your nervous system."[15] MacKay reasons that "conscious experience is *embodied in* our brain activity: neither on the one hand identical with it, nor on the other hand quasi-physically interactive with it."[15] For MacKay, to copy the brain to a computer would be to create a correlation, not a translation.

Even if mind uploading is never attempted, the belief that, given sufficient advances in technology, in principle it could be done has subtle implications for neuroethics now. The view that the human mind is equivalent to a computer is possible, as C. Ben Mitchell and colleagues have argued, "only on the assumptions of the scientific materialist, which reduce human persons to their biological parts and biotechnological enhancements."[16] In the overvaluing of computer technology, proponents of mind uploading undervalue human dignity.

There is still a great deal about how the brain works that neuroscience has not deciphered. Although neuroscience has shed considerable light on the functions of the brain, it lacks the ability to explain the phenomena of consciousness, personal agency, conscience, moral responsibility, the continuity of identity over time, or human purpose. Of these, consciousness seems the most elusive, if not irreducibly subjective. If consciousness were reducible to neural activity, then, writes Jaki, "it should be subject, like any other physical process, to cybernetical analysis. But evidently, the phenomenon of consciousness slips through the sieves of cybernetics no less swiftly than it keeps eluding the anatomist's scalpel or the neurophysiologist's electrodes."[17]

Until such time as these unknowns can be explained in scientific terms, which seems doubtful because they transcend materialistic descriptions, projects intended to upload a human mind to a computer would risk leaving behind essential aspects of what makes one human.

A One-Way Port

Before departing one's body and—assuming for the sake of argument that it is even possible—uploading one's mind to the realm of cyberspace, one ought to pause to consider what kind of world that might become for those who choose to dwell within it. Once uploaded, there is no going back. In a reversal of the choice faced by Lot's wife, going forward into the realm of mind uploading would turn one's fleshly brain to solid matter.

At a recent conference at Oxford University, Bostrom told his academic audience, "I personally believe that once human equivalence is reached, it will not

be long before machines become superintelligent . . . our future is likely to be shaped by them, for the better or the worse. Superintelligence . . . could be an extremely powerful ally that could help us solve a number of other problems that we face," but he added that superintelligence could also be "extremely dangerous," even to threaten our extinction.[18]

In his novel *The Transhumanist Wager*, Zoltan Istvan's transhumanist protagonist argues,

> Our biology severely limits us. . . . The transhumanist believes we should immediately work to improve ourselves via enhancing the human body and eliminating its weak points. This means ridding ourselves of flesh and bones, and upgrading to new cybernetic tissues, alloys, and other synthetic materials, including ones that make us cyborglike and robotic. It also means further merging the human brain with the microchip and the impending digital frontier. Biology is for beasts, not future transhumanists. . . . If you're not necessary and do not serve a transhuman purpose, and you also destroy resources for those who are necessary and serve transhuman purposes, you may not be allowed to exist.[19]

There is no guarantee that existing as computational entities liberated from organic bodies would ensure autonomy or happiness, and every reason to think not. Posthuman entities existing as data clouds might find themselves at the mercy of vastly more powerful computational forces. Distinctions between uploaded human intelligences and artificial intelligences might disappear as both further evolve and contend for resources. Bereft of human programmers, such a world might know no ethical boundaries. Absent human persons, one might also ask whether that would matter.

One More Experiment

Arthur C. Clark imagined a possible future of artificial intelligence in the 1968 MGM movie *2001: A Space Odyssey*. In a now famous line, the film's antagonist, a sentient computer named HAL 9000, takes life-threatening action against astronaut Dave Bowman when he attempts to shut it down. Bowman asks, "Open the pod bay doors, HAL." HAL calmly replies, "I'm sorry, Dave. I'm afraid I can't do that."

When I did my own experiment using Apple's voice recognition program that is built into the iPhone, life imitated art. With playful intent, I spoke into my phone, "Open the pod bay doors, Siri." Without a blink in its screen, the phone replied, "Sorry, William. I don't do pod bay doors."

The iPhone was, of course, only generating text according to its programming, treating words not as parcels of meaning to convey intent but as bits of neutral data to be shuffled in the mechanical process of input and output. My iPhone was not actually thinking. Or was it?

References

1. Meredith Bennett-Smith. Stephen Hawking: Brains Could Be Copied to Computers to Allow Life after Death. *The Huffington Post*, September 24, 2013. Accessed at: http://www.huffingtonpost.com/2013/09/24/stephen-hawking-brains-copied-life-after-death_n_3977682.html
2. Spohrer J. NBICS (Nano-Bio-Info-Cogno-Socio) convergence to improve human performance:

Opportunities and challenges. In: Roco MC and Bainbridge WS, editors, *Converting Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science*. Arlington, Virginia: National Science Foundation, 2002, p. 110.

3. *Three Pounds*, produced by Peter Ocko and Ken Topolsky, CBS, 2006.
4. Kurzweil R. *The Singularity is Near: When Humans Transcend Biology*. New York: Viking, 2005, pp. 145.
5. Kurzweil, p. 199.
6. Kurzweil, p. 377.
7. Bostrom N. *Ethics for Intelligent Machines: A Proposal*, 2001, cited in Kurzweil, p. 369.
8. Cheshire WP. Glimpsing the grey marble. *Ethics & Medicine* 2007; 23(2): 119-121.
9. Kurzweil, p. 124.
10. Perea G, Sur M, Araque A. Neuron-glia networks: integral gear of brain function, *Front Cell Neurosci* 2014; 8: 378.
11. Jaki SL. *Brain, Mind and Computers*. South Bend, IN: Gateway, 1969, pp. 75-80.
12. Pearcey N. *Finding Truth*. Colorado Springs, CO: David C. Cook, 2015.
13. Kurzweil, p. 200.
14. Kurzweil, p. 384.
15. MacKay DM. *Behind the Eye* (Gifford Lectures delivered in 1986). Oxford, UK: Basil Blackwell, 1991, pp. 1-11, 262-266.
16. Mitchell CB, Pellegrino ED, Elshtain JB, Kilner JF, Rae SB. *Biotechnology and the Human Good*. Washington, DC: Georgetown University Press, 2007, p. 42.
17. Jaki, p. 128.
18. Shead S. Future of humanity: Nick Bostrom supports Stephen Hawking's AI predictions. *Techworld*, May 18, 2015. Accessed at: http://www.techworld.com/news/operating-systems/ai-guru-nick-bostrom-supports-stephen-hawkings-rising-robot-predictions-3612024/
19. Istvan Z. *The Transhumanist Wager*. Reno, Nevada: Futurity Imagine Media, 2013, p. 278.

William P. Cheshire, Jr., MD, MA, is Professor of Neurology and Chair of the Ethics Committee at the Mayo Clinic in Jacksonville, Florida; Chair of the Ethics Committee for the Christian Medical and Dental Associations; and Senior Fellow at The Center for Bioethics & Human Dignity. The views expressed herein are his own and do not necessarily reflect the positions of the professional organizations with which he is affiliated. He currently resides in Ponte Vedra Beach, Florida, USA.