Journal of Evolution & Technology

A peer-reviewed electronic journal published by the Institute for Ethics and Emerging Technologies

ISSN 1541-0099

22(1) –December 2011

Human Enhancement and the Computational Metaphor

James Ogilvy
Global Business Network
Email address: jaogilvy@gmail.com

Journal of Evolution and Technology - Vol. 22 Issue 1 – December 2011 - pgs 81-96

Abstract

This paper affirms human enhancement in principle, but questions the inordinate attention paid to two particular forms of enhancement: life extension and raising IQ. The argument is not about whether these enhancements are possible or not; instead, I question the aspirations behind the denial of death and the stress on one particular type of intelligence: the logico-analytic. Death is a form of finitude, and finitude is a crucially defining part of human life. As for intelligence, Howard Gardner and Daniel Goleman show us the importance of multiple intelligences. After clarifying the notion of different psychological types, the paper takes five specimens of a distinct type and then studies the traits of that type through their examples. Seeking a pattern connecting those traits, the paper finds them bound together by the embrace of the computational metaphor for human cognition and then argues that the computational metaphor does not do a good job of describing human intelligence. Enlisting the works of Jaron Lanier and Ellen Ullman, the paper ends with a caution against pushing human intelligence toward machine intelligence, and points toward the human potential movement as a possible ally and wise guide for the transhumanist movement.

The argument for human enhancement is at once simple and complex: very simply, to quote Zarathustra, "What is great in man is that he is a bridge and not an end" (Nietzsche 1954, 127). Enhancement is paradoxically the essence of creatures whose existence precedes their essence – the formula, if you will, for existentialism. But once the door toward enhancement has been so simply opened, questions about which enhancements, to whom, and in what degree, become fantastically complex.

This essay makes short shrift of essentialist objections to any and all enhancement. Most of the essay is then devoted to the question: what's wrong with the inordinate amount of attention paid to life extension and IQ?

Short shrift

Let's not waste time with straw men. Shoes enhance the soles of your feet; clothes enhance the ability of your skin and hair to insulate your body; glasses enhance vision. We're not about to go naked into the world, and once we acknowledge Nietzsche's "bridge" of enhancement as the slippery slope we stepped out onto long ago, we're not going to outlaw lasik surgery because we've fixed on spectacles as part of an ahistorical human nature.

That said, there are distinctions to be drawn between lasik surgery available to anyone and life extension available to only a few. Likewise there are distinctions to be drawn between making sure that all God's chillun' got shoes, and smart pills for an affluent few.

Surely there are arguments that can be made against the injustice of unequal opportunities for enhancement, and I don't want to make light of such inequities. But in this essay I am less concerned with issues of inequity than in how we make choices about what to enhance. Once we see humanity as a "bridge", once we take seriously our existential historicity, then we're bound to embrace a certain *elongation* of the human condition. That is to say, some will always be further along than others. The good news of our capacity for progress comes with the potentially bad news of an inevitable unevenness of achievement.

I don't want to dismiss issues of inequality and inequity. They are real. But the focus of this essay is elsewhere: I want to push back against the emphasis on (a) life extension unto immortality, and (b) the enhancement of IQ. Both life expectancy and IQ are eminently quantifiable. I would rather focus on quality of life. A longer life is not necessarily a better life. Nor is a smarter person necessarily a wiser person.

Finitude and the lure of life extension

Consider life extension. While increases in life expectancy due to improved nutrition and healthcare are certainly to be applauded, the prospect of immortality constitutes a violation of our finitude, and this, I will argue, is a bad thing. The point is not simply that we don't want to prolong decrepitude. Ray Kurzweil (2005) is smart enough to see that we don't want to feel like centenarians for many centuries; he wants to feel forty forever. That's why he takes over 200 pills a day.

What is wrong with aspirations to immortality? I'll offer two arguments, one for the individual and a second for the species. For the individual, the aspiration to immortality, the denial of death, is the denial of a part of life that gives it definition. Our finitude is not just a number like "fourscore and ten"; it is a constraint that makes for meaning and significance. Heideggerian being-towards-death isn't so much a brooding about nothingness; it is better understood as the embrace of somethingness rather than anythingness. You can be a butcher *or* a baker *or* a candlestick maker . . . but not all three and anything else besides.

The decisiveness of Kierkegaard's *Either/Or* is closer to the existential significance of finitude than the fact that we all die sooner or later. While the affect of Heidegger's staring into the abyss might make it look as though life were being enhanced by a simple contrast with death-whose-finality-is-fully-absorbed, I think that this is the wrong lesson to take away from Heidegger. Once he turns from the Teutonic weightiness of *Being and Time* toward his later lectures on Nietzsche, we see a lighter embrace of *amor fati*, a "love of fate" that comes with the doctrine of eternal return. To be able to say of one's life, "Thus it was and so I willed it, and so would I will it again

and again for all eternity . . ." – this is not the denial of death but the affirmation of a particular, unique, and finite life. Just so.

Surely it could be argued that the constraint and definition of a life can be achieved in ways other than snipping it short after a finite number of years. Yes, the definition of different types and traits would be possible for immortal beings, just as it was for the gods and goddesses in polytheistic pantheons. But the denial of death has a further downside for the species, one that can be captured by observing that heaven has no history.

There is no such thing as progress in heaven. The problem with saying, "We are as gods and might as well get good at it" (Brand 1968) is not simply hubris. Rather, in taking the gods as models for an enhanced humanity, we would deny our historicity and with it the temporality that is a significant part of our humanity. Imagine life in a society of immortals. Where are there opportunities for improvement? Selection by way of death is evolution's way of culling old models to make room for new and improved models. Obsolescence, whether planned or unplanned, is a precondition for the possibility of novelty. Gods and goddesses may have little use for novelty, but we do. Our species needs death in order to evolve, in order to be a bridge and not an end. Mortality is a feature, not a bug. Without it, we, as a species, would freeze.

For the "individual existing human being" (a favorite phrase of Kierkegaard's), it may be too much to expect a willingness to "step aside" for the sake of the improvement of the species. You don't see many professors giving up tenure for the sake of making room for new blood in their departments. Nor are political incumbents noted for their willingness to give up their offices. But for that very reason – the expectable self-interest of incumbency – we who take the long view toward the interests of the species and not just our own sweet selves should be especially cautious about supporting programs and policies that indulge the denial of death. Tenure is a mixed blessing.

Intelligence. What?

What about the enhancement of IQ, whether by "smart pills" or other means? Here again I'm less concerned with the means, or with issues of equity, than with the aspiration in the first place. Yes, it would be nice to improve one's memory and one's ability to solve tough problems. Intelligence is a good thing. But what *is* intelligence? More precisely, is there an important difference between *human* intelligence and the computational power of machines that, according to Kurzweil and others, will be able to match human intelligence by 2030 or so?

Like love and consciousness, intelligence is famously hard to define. But just as the intractability of love has not discouraged poets from counting the ways we love one another, so psychologists have not been shy about trying to pin down intelligence. Early approaches by Francis Galton, Alfred Binet and Lewis Terman assumed that intelligence was a single capacity susceptible to monolinear measurement, and measure it they did with instruments like the famous IQ test. Because there were obvious problems involved in measuring something so ill-defined, tests for intelligence eventually served as operational definitions for what it was they were testing. As Harvard psychologist, E. G. Boring (1923) put it, "Intelligence *is* what the tests test." But the uses to which such tests are put led to criticisms: that the tests were biased in favor of those with a certain culture or background; that the tests ignored certain capacities that could or should be regarded as part of intelligence. Herrnstein and Murray's influential book, *The Bell Curve* (1994), provoked a torrent of criticism for arguing that intelligence is a single property distributed differently among different races.

Rather than claiming that intelligence is one capacity that can be measured on a single scale, Howard Gardner (1983, 1985, 1999) has built a career around the exploration of *multiple intelligences*. In his early work he lists seven: (1) linguistic, (2) logical-mathematical, (3) musical, (4) bodily-kinesthetic, (5) spatial, (6) interpersonal, and (7) intrapersonal. In his later work Gardner adds to his list (8) naturalist (or ecological) intelligence, but balks at the addition of spiritual intelligence, or something he calls "existential intelligence," or moral intelligence. Building on and adding to Gardner's work, Daniel Goleman published his immensely successful *Emotional Intelligence* (1995). Can this list of multiple intelligences be further extended? Is a sense of humor a form of intelligence? Some have alleged that it should be, but that Gardner lacks a funny bone.

Apart from quibbles about the length or composition of the list of multiple intelligences, what has become abundantly clear is that the capacity for logico-analytic calculation is but a small component of what many of us would want to count as intelligence. Computational capacity, whether measured in terms of storage capacity or speed, is but a part of what counts as intelligence. There are as many ways of being smart as there are ways of being stupid. In academia, we are all familiar with people my wife would call, "smart in books but dumb in life." As David Brooks argues in *The Social Animal* (2011), success at assembling the several parts of a good life – economic success, happiness, and the maintenance of strong relationships – does not correlate that well with purely cognitive capacity.

So when Ray Kurzweil tries to put a date on when computers will "exceed human intelligence," the question we should be considering is not so much about the speed or capacity of computers; the question we should be considering is whether Kurzweil has a proper understanding of just what *human* intelligence *is*. I have no doubt that computers already exceed my capacity for a range of tasks from finding cube roots to remembering names and dates. But when it comes to good judgment, or what the Greeks called *phronesis* (practical wisdom), neither the progress in AI (artificial intelligence) to date, nor its prospects for the future, invite optimism. Computers are great at calculation, but they lack common sense.

Questioning of aspirations

With respect to aspirations to immortality on the one hand, and the enhancement of IQ on the other, the question is not so much whether we can, but why would we want to? Yes, of course, an added increment to length of life or IQ looks like a good thing in particular cases, but a generalized obsession – as manifested, say, in the consumption of over 200 pills a day – may lead to too much of a good thing. It strikes me as one-sided, intemperate, almost pathological.

Now this is a lot to claim – not that Kurzweil and his transhumanist colleagues are so much *incorrect* in their estimates for a competition between computers and humans, but rather that they are somehow sick for even wanting to go there. The argument ceases to be logical and evidential, and swerves dangerously toward the *ad hominem*. But that is not exactly the nature of the argument I want to make. I don't want to accuse transhumanists of being crazy in some psychologically reductionist and dismissive way. But I do want to identify a syndrome, a type if not a stereotype, an historically emergent combination of cultural, psychological and technological factors that are coming together in ways that are both pathological and highly functional.

To say that I am suggesting some new category for the next edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM IV 2004) would be to take a stance that is too narrowly psychological, and too tilted toward the pathological. I prefer an approach that is less

reductive, more holistic, and more in the tradition of critical theory but without the Marxism. Let me explain with examples.

Psychographics, history, and character types

After the Second World War, in the wake of consternation at how a literate and civilized people like the Germans could plunge so tragically toward the atrocities wrought by Fascism, Theodor Adorno and some of his colleagues undertook a research program with linked objectives: (1) to identify the character type that is susceptible to fascism; and (2) understand its etiology in sufficient detail so as to assure that such atrocities would happen *never again.* Given the recent history, it was apparent to them that they could not simply ask the man on the street, "Are you a Fascist?" Instead they identified a series of seemingly innocent and seemingly unrelated questions and corresponding answers that were in fact highly correlated with Fascistic behavior. They published the results of their work in an impressive tomb entitled *The Authoritarian Personality* (Adorno et al. [1950] 1969). Meanwhile renegade psychologist Wilhelm Reich stepped out of the narrow confines of Freudian psychology to fetch a wider cultural-historical compass in his equally weighty book, *The Mass Psychology of Fascism* (Reich [1946] 1970). Flawed as they may have been – Reich ended his life certifiably insane – both of these works share the merit of seeing the human condition whole. Both works respect the incredible complexity of both nature and nurture; both weave rich tapestries combining instinctual dynamics from the (subjective) inside with cultural and historical conditions on the (objective) outside. Neither narrowly psychologistic, nor reductionistic toward cultural or economic determinism, both works take the intellectual high road toward an ambitious comprehension of social, psychological, cultural, historical and economic factors all wrapped up in the syndrome known as the authoritarian personality.

And they are not alone. Others have acknowledged that psychopathologies have histories – that different decades in different cultures have their sickness *du jour.* For Freud's Vienna it was hysteria, whose etiology bore traces of the outside influence of a repressive Victorian culture and morality. During the 1950s and 60s, paranoid schizophrenia gained in currency as a presenting symptom. Erik Erikson (1968) coined the term "identity crisis" to capture a syndrome he identified among WWII vets whose symptoms would later by classified as PTSD (post-traumatic stress disorder). Here again the unique historical sources of the "stress" represent outside influences quite independent of the psychic drives and instincts studied by a narrower Freudian psychology.

During the 1970s and 80s the fashionable pathology became the narcissistic borderline personality disorder, whose portrait was painted not only by clinical psychologists like Heinz Kohut (2000), but also by sociologists like Robert Bellah (1985, 1991), Philip Slater (1990), and Christopher Lasch (1978). Here again we were asked to acknowledge not just the psychic inner workings of some ahistorical drives and instincts, but the historico-cultural influences of a uniquely American hyper-individualism.

Even as these several "outside-in" approaches to the psyche from the broader realms of culture, history and economics were enhancing our understanding of the history of madness, humanistic psychology and the human potentials movement were giving us a psychology of health rather than sickness. In the works of Abraham Maslow (1968), Carl Rogers (1989) and Michael Murphy (1992), we find an appreciation for the higher reaches of human performance. In the words of R. D. Laing (1965), a psychologist who was also sensitive to broader social and cultural dynamics, sometimes what psychologists see as a "breakdown" may in fact be a "breakthrough." Likewise

Kohut was quick to acknowledge that many of the so-called narcissists he treated were very successful, highly performing people.

The psycho-social typology game is not just about *ad hominem* reduction and dismissal. It is instead a hermeneutic: a means of interpreting various markings as symptoms of various traits. Its varieties are legion, from astrology (Tarnas 2006) to the Enneagram (Riso and Hudson 2000). I, for one, find almost nothing useful in these latter two interpretive systems. But I understand the impulse behind them: there really are *different types* of people, and wouldn't it be nice to have a tidy system for classifying them all, and doing so respectfully, without seeing each as a species of pathology.

Where astrology and the Enneagram are, in my opinion, lacking in empirical validation, other typologies have been subject to rigorous confirmation. For several years in the 1980s I served as director of research for the Values and Lifestyles (VALS) Program at SRI International (formerly Stanford Research Institute). Arnold Mitchell (1986) invented a typology based on a creative synthesis of Maslow's famous hierarchy of needs bisected by David Riesmann's (1950) distinction between the "inner-directed" and the "other-directed." Mitchell's typology contained nine different lifestyles that differed not simply by demographic characteristics of age, income and education level, but also by the different values espoused by members of very similar demographic segments. Schooled in the kind of public opinion survey techniques initiated by Adorno et al. in their research on the authoritarian personality, we in the VALS Program conducted national probability sample surveys of Americans that were then rigorously correlated with purchase patterns as tracked by Simmons Market Research. Using the results of this research, advertisers and marketers were better able to target the right consumer with the right message about the right product through the right media channel. To cite just one example that will show just how wrong things can go when there's a mismatch between economic supply and psychodynamic demand, consider an ad for 0% financing for Chevy trucks that was broadcast during a break in a rerun of *Brideshead Revisited.* Yes, this stuff works, as is most clear when it doesn't.

Labels for other psycho-social types – characterized not only by their psychology but also their socio-economic context – will be familiar, viz. Yuppies (young urban professionals exuding a get-rich-quick ambition); or David Brooks' (2000) construct, the Bobos (bourgeois bohemians – aging hippies who make some money and then take on the trappings of suburban family life).

Now it is in the spirit of this intellectually rich and lively tradition of psycho-social typologizing that I would like to identify the traits of a new type: the noble nerd. We've all heard the usually derogatory term, "nerd." I add the modifier "noble" with the explicit intention of *not* wanting to pathologize this new type. Even as I would acknowledge that pathologies are possible – even that we know them chiefly by their pathologies – still, like Kohut and his narcissists, or Brooks and his Bobos, I want to respect both the good intentions and high achievements of many a noble nerd. Nonetheless, I would have you see them *as one type among other types,* not as models or paradigms for the human condition. There is a syndrome that has both its positive and negative characteristics. As with some of the other psycho-social types we have just surveyed, noble nerds represent a complex coming together of social, psychological, cultural, historical and economic influences at a particular juncture in the world problematique. Precursors can be found in prior eras. Successors will modify the type. But here and now, in places like Silicon Valley, Boston's Rte. 128, Raleigh's Research Triangle, and Austin, Texas, this type runs rampant.

Getting to know the noble nerds

Ye shall know them by their Facebook pages, their LinkedIn profiles and their Twitter feeds. So-called "social networking" is their massive compensation for their fundamentally asocial instincts (they are heirs to Lasch's *Culture of Narcissism*). But how much better that they reach out and at least click someone rather than receding into the solitary existence of Japan's *hikikomori*. Who? In today's Japan there are more than a million young people, mostly male, who will not leave their bedrooms. According to Michael Zielenziger, who parlayed his eight years as Reuters bureau chief in Tokyo into a book on the subject (2006), these young people are not clinically crazy. If air-lifted out of Japan to Hong Kong, Vancouver or New York, they often thrive. Nor are they the victims of some sort of digital addiction to their computer screens. Yes, as Sherry Turkle, has informed us, too much *Life on the Screen* (1995) *can* be dangerous to your mental health. But for the *hikikomori*, their withdrawal seems to have more to do with a mismatch between their personal aspirations as twenty-first-century globally informed youth on the one hand, and on the other the social constraints and expectations imposed by a very strong and homogeneous Japanese culture. *Hikikomori* is not a good candidate for DSM V. This new type is not just a new species of psychological pathology. The *hikikomori*, I would argue, represent a particularly unhappy version of noble nerds. Their experience may be intensely psychological, and for the most part painful, but its etiology has a lot to do with a unique juncture of social, cultural, technological and historical circumstances.

To get a better sense of the complex dynamics of the noble nerd, some case studies are in order. Here the method is that of the naturalist. Go out and trap a few that you think are representative. Put them together in a display case and examine them to see whether they have any common traits. By an inductive process of generalization from particulars, try to identify some common themes. Weave those themes into a coherent tapestry of mutual affinities (much looser than necessary entailments), and then test the results by making some deductive inferences that are falsifiable.

This method is eminently subject to errors. In investigating what she takes to be *fish*, if the naturalist nets some dolphins and/or turtles along with all of the salmon, cod and haddock, some odd inductive inferences will follow. Features that characterize some instances of the class may not apply to others. We are creating a construct here, not discovering a natural kind. Its dimensions are less defined by natural evolution than by historical and cultural change. Its boundaries are therefore fuzzy and in flux. Its validity should be measured less by its faithful correspondence to some objective silhouette and more by its usefulness in generating insights.

With that much of methodological caveats out of the way, let me boldly state that my candidates for the display case include not only Ray Kurzweil but also Herman Kahn, Marvin Minsky, Steven Wolfram, and Ed Fredkin. Is there a "dolphin" in this mix, someone who looks like a member of the class but isn't? Perhaps. Mistakes will be culled in further iterations of the method, but for now we will look to these men as paradigm cases of the noble nerd. Each one is widely regarded as a genius. "Criminally intelligent" is a phrase that sits easily on Herman Kahn, whom many regarded as a moral monster for his ability to joke about mega-deaths in his classic, *On Thermonuclear War* (1960).

But what kind of "intelligence" is this that is obsessed with gaining *more* life and *more* intelligence? In deriving a profile based on these *Übermenschen*, I hope to tell a tale that is cautionary rather than aspirational for the transhumanist movement. Beware of what you want. So-called human enhancement may not be all that you hoped it would be. And/or, it may be more of some things you should fear.

So what are some of the "traits" of the noble nerd? Let's begin with the most obvious: smart, socially inept, jocular, ambitious, numerate . . . you know the type *as* a type. But to understand the syndrome better, let's drill down on some examples and then see if we can connect these dots to some other less obvious dots.

Smart: Each of our specimens has been often and publicly heralded as brilliant. Wolfram won a MacArthur "genius" award. Friends of mine who know Minsky joke that if one were able to quantify just how much knowledge there is in the world, the unit of measure would be in Minskies. When Kahn was given an IQ test in the course of joining the military, he got the highest score ever recorded by his testers. Ed Fredkin, at 34, became a full professor at MIT.

Socially inept: Ed Fredkin:

> I couldn't conduct a conversation with a girl or arrange for a date or get invited to a dance. I was not invited to a single party or dance throughout my whole high school time. Not once. . . . [K]ids would be choosing sides for a game of something – it could be touch football. They'd choose everybody but me and then there'd be a fight as to whether one side would have to take me. One side would say, "We have eight and you have seven," and the other side would say, "That's okay." They'd be willing to play with seven. (Wright 1998, 16f.)

Herman Kahn was a great conversationalist and very gregarious . . . but his was not an ordinary way of interacting. "At least as important as adulation was condemnation. He really wanted to be cursed and damned. He just gloried in it" (Ghamari-Tabrizi 2005, 82). Read chapter 3 of Ghamari-Tabrizi's remarkable book on Kahn, *The Real Dr. Strangelove*, and you be the judge.

Jocular: Kahn's humor offended many. How could he crack jokes about mega-deaths? Kurzweil concludes many chapters in *The Singularity is Near* with fanciful and amusing "conversations" among the likes of Charles Darwin, Ned Ludd and Tim Leary. Fredkin, said a colleague, "doesn't really work. He sort of fiddles."

> "Very often he has these great ideas and then does not have the discipline to cultivate the ideas" "There is a gap between the quality of the original ideas and what follows. There's an imbalance there." Fredkin is aware of his reputation. In self-parody, he once brought a cartoon to John Macone's attention. In it, a beaver and another forest animal are contemplating an immense man-made dam. The beaver is saying something like, "No, I didn't actually build it. But it's based on an idea of mine." (Wright 1998, 35)

Ambitious: Wolfram's ambition bleeds from the very title of his book, *A New Kind of Science*; its size – over a thousand pages; and the sheer number of disciplines he attacks. And he has made a lot of money from his software, Mathematica. Likewise, Kurzweil writes very big books on just about everything under the sun, and makes a very good living from the several companies he has built. Fredkin bought an entire island in the Caribbean with the fruits of his brilliance. He wanted to develop an institute that would use artificial intelligence to disambiguate the language of diplomacy. "This particular brainstorm was but one reflection of Fredkin's impulse to save the world" (Wright 1998, 45). And Kahn, too, was out for nothing less: "We take God's view. The President's view. Big. Aerial. Global. Galactic. Ethereal. Spatial. Overall. Megalomania is the standard occupational hazard" (Ghamari-Tabrizi 2005, 70).

Numerate: I choose this little used word carefully in order to point toward a worldview that is as metaphysical as it is mathematical. It's not simply that these men like to count and calculate. It's rather that they believe that absolutely everything is denumerable. And, consistent with their ambition, they like to traffic in very big numbers, exponentials, ten to the nth where *n* is often a double or triple digit. Look at all the logarithmic plots in *The Singularity is Near*. Listen to Kahn carry on about the difference between 40 million and 80 million deaths following a nuclear exchange. Marvel at the sheer number of neurons in Minsky's *Society of Mind* (1985), whose basic argument is that intelligence will result if you correctly wire together *enough* stupid agents. For Wolfram, whose arguments often depend on the emergent outcomes of thousands of iterations of his cellular automata, "More is Different," to quote the title of Phil Anderson's famous article (1972). And finally, for Fredkin . . . I'll let science writer, Robert Wright, describe the situation better than I can, but first you need to know that Fredkin believes that the universe is not just *like* a computer. "Ed Fredkin thinks the universe is a computer. A really big one" (Wright 1998, 4).

> Fredkin has an interesting way of expressing his insistence that all physical quantities be rational. (A rational number is a number that can be expressed as a fraction – as a *ratio* of one integer to another. In decimal form, a rational number will either end, like 5/2 in the form of 2.5, or repeat itself endlessly, like 1/7 in the form of 0.142857142857142 . . .) He says he finds it hard to believe that a finite volume of space could *contain* an infinite amount of information. It is almost as if he views each parcel of space as having the digits describing it actually crammed into it. This seems an odd perspective, one that confuses the thing itself with the information representing it. But such an inversion between the realm of things and the realm of representation is common among those who work at the interface between physics and computer science. Contemplating the essence of information seems to affect the way you think. . . "I've come to the conclusion," he says, that the most concrete thing in the world is information." (Wright 1998, 26-27)

So much for this quick tour through selected evidence for the first five and most obvious traits of the noble nerds. Now for the less obvious. How do these dots fit together? Is there a pattern that connects? I believe that there is and that Fredkin, the least famous of our five specimens but maybe the most extreme exemplar of the type, holds the key. His somewhat bizarre belief that the universe *is* a computer goes one step further, a step toward literalism, in what others speak of as "the computational metaphor."

The computational metaphor and its consequences

Marvin Minsky is among the earliest advocates of so-called "strong AI," the belief that the brain functions like a computer and that once we have enough computational power, we will achieve artificial intelligence. Both Wolfram and Kurzweil are clearly committed to the view that the human brain works like a computer. We haven't achieved artificial intelligence *yet* because, on the one hand, we don't have enough computing power, and on the other, we don't really understand how the brain works. But we will. It's just a matter of time before a computer will pass the Turing test. Kurzweil is convinced – and has placed bets to back his conviction – that computers will surpass human intelligence sometime before 2030.

But again, what *is* intelligence? Whether or not you acknowledge all of the types of intelligence described by Howard Gardner and Daniel Goleman, it's possible to question whether digital computation provides a good metaphor, much less a rigorous model, for human intelligence. Among the strongest critics of the computational metaphor for the brain are two Berkeley philosophers, Hubert Dreyfus and John Searle. In 1972, Dreyfus published *What Computers*

Can't Do: The Limits of Artificial Intelligence. Given that Minsky was a real scientist and Dreyfus a mere philosopher, Dreyfus was, at the time, the contrarian considered to be in the minority. Nonetheless, his book kicked up enough of a storm that he felt called upon to publish a revised edition in 1979. Then, in 1992, he published *What Computers Still Can't Do,* in which he writes:

> Almost half a century ago computer pioneer Alan Turing suggested that a high-speed digital computer, programmed with rules and facts, might exhibit intelligent behavior. Thus was born the field later called artificial intelligence (AI). After fifty years of effort, however, it is now clear to all but a few diehards that this attempt to produce general intelligence has failed. This failure does not mean that this sort of AI is impossible; no one has been able to come up with such a negative proof. Rather, it has turned out that, for the time being at least, the research program based on the assumption that human beings produce intelligence using facts and rules has reached a dead end, and there is no reason to think it could ever succeed. (Dreyfus 1992, ix*)*

Dreyfus' colleague, John Searle, has also taken aim at the computational metaphor for consciousness, first in a famous article (Searle 1980) and then in a series of books (1984, 1992, 1997, 2002, 2007). His critique centers around his so-called Chinese Room argument. He asks us to imagine a person, possibly Searle himself, who is ignorant of Chinese sitting in a closed room:

> Imagine that you carry out the steps in a program for answering questions in a language you do not understand. I do not understand Chinese, so I imagine that I am locked in a room with a lot of boxes of Chinese symbols (the database), I get small bunches of Chinese symbols passed to me (questions in Chinese), and I look up in a rule book (the program) what I am supposed to do. I perform certain operations on the symbols in accordance with the rules (that is, I carry out the steps in the program) and give back small bunches of symbols (answers to the questions) to those outside the room. I am the computer implementing a program for answering questions in Chinese And this is the point: *if I do not understand Chinese solely on the basis of implementing a computer program for understanding Chinese, then neither does any other digital computer solely on that basis, because no digital computer has anything I do not have.* (Searle 1997, 11)

Searle's thought experiment serves to illustrate the following simple logic:

1. Programs are entirely syntactical.
2. Minds have a semantics.
3. Syntax is not the same as, nor by itself sufficient for, semantics. (Searle 1997, 11)

I think Searle's argument holds water, as do Dreyfus's arguments in *What Computers Still Can't Do,* but the fact of the matter is that there have been over a hundred published attacks on Searle's argument, including a sustained argument in *The Singularity is Near* (Kurzweil 2005, 458-69). This debate is not over, the case has not been settled. Despite confident declarations of the death of strong AI by Dreyfus and Searle, the computational metaphor is alive and well. See, for example, how the founding editor of *Wired* magazine, Kevin Kelly, explicitly embraces the computational metaphor for mind in his recent book, *What Technology Wants*:

> We tend to interpret the mysteries surrounding life in imagery suggested by the most complex system we are aware of at the time. Once nature was described as a body, then as a clock in the age of clocks, then a machine in the industrial age. Now, in the "digital age," we apply the computational metaphor. To explain how our minds work, or how

> evolution advances, we apply the pattern of a very large software program processing bits of information. None of these historical metaphors is wrong; they are just incomplete. Ditto for our newest metaphor of information and computation. (Kelly 2010, 63-64)

And Kelly is not alone. Though I don't have poll data to prove it, I would guess that most noble nerds embrace the computational metaphor for mind. But there are consequences to this embrace. For one thing, it becomes very difficult to defend free will if you think that the brain is a computer. Like some others, Kelly tries to rescue freedom by appeal to quantum indeterminacy (Kelly 2010, 307-311). But the freedom so gained is the freedom of a leaf in the wind, a negative lack of constraints rather than a positive choice of purposes for reasons rather than causes. No less a philosopher than Daniel Dennett (1984) gives a definitive refutation of the argument from indeterminacy.

So another consequence of embracing the computational metaphor, and lacking an adequate defense of free will, is to simply bite the bullet of determinism, which is what both Wolfram and Kurzweil end up doing (Kurzweil 2005, 521f n. 65).

Yet another consequence of embracing the computational metaphor – and here we come to the main point of this paper – is this: Rather than claiming that machines can rise to an equivalence with or supersede human intelligence, what the noble nerds are ultimately doing is lowering human intelligence to its most mechanistic aspects.

Listen closely to Kurzweil:

> The Singularity will represent the culmination of the merger of our biological thinking and existence with our technology, resulting in a world that is still human but that transcends our biological roots. There will be no distinction, post-Singularity, between human and machine. (Kurzweil 2005, 9)

Fredkin, per usual, is even more explicit and extreme:

> In the early 1980s, Fredkin – tired, presumably, of beating around the bush – taught a course at MIT on saving the world. The idea was to view the world as a giant computer and to write a program that, if methodically executed, would lead to peace and harmony – the "global algorithm," it was called. Along the way, an international police force would be formed and nations would surrender some of their autonomy to international tribunals. "It's a utopian idea," Fredkin concedes, but he adds with emphasis that it's not anything so simplistic as a formula for *instant* utopia. "This is a *series of steps* . . . that gets you to utopia." If more people would take the plan seriously, he says, it could succeed. "I'll make this strange sort of arrogant statement that the reason people think my ideas aren't practical is that . . . they don't understand that if they would just sort of act like machines it would all work." (Wright 1998, 45-46)

Precisely: If we'd all just act like machines . . . then surely we'll lose out in the competition with machines that are getting ever bigger and faster. Then computer "intelligence" will surpass human intelligence for sure.

The question, then, is not about the rate of advances in computing power. The question is the nature of genuinely human intelligence. And the answer to this question, according to Dreyfus, Searle, and even more persuasively, Berkeley anthropologist Terrence Deacon (1997, 2011

forthcoming), is that human intelligence is not at all like that of a computer, however grand the dreams and/or the achievements of the AI research program.

In relying on faster computers as the means to enhance humanity, we run the risk of actually degrading humanity. This is a large part of the point of Jaron Lanier's important book, *You are not a Gadget*:

> People can make themselves believe in all sorts of fictitious beings, but when those beings are perceived as inhabiting the software tools through which we live our lives, we have to change ourselves in unfortunate ways in order to support our fantasies. We make ourselves dull. (Lanier 2010, 156-57)

What kind of enhancement is this?

Kurzweil is intriguingly ambidextrous in the way he both conflates and distinguishes "nonbiological intelligence" and human intelligence. On one page he writes:

> Once nonbiological intelligence gets a foothold in the human brain (this has already started with computerized neural implants), the machine intelligence in our brains will grow exponentially (as it has been doing all along), at least doubling in power each year. In contrast, biological intelligence is effectively of fixed capacity. Thus, the nonbiological portion of our intelligence will ultimately predominate. (Kurzweil 2005, 28)

And then on the very next page:

> Ultimately, the entire universe will become saturated with our intelligence. This is the destiny of the universe . . . We will determine our own fate rather than have it determined by the current "dumb," simple, machinelike forces that rule celestial mechanics. (Kurzweil 2005, 29)

The universe will become saturated with *our* intelligence? A genuinely human intelligence? Or won't it rather be saturated with the intelligence of what Kelly (2010) calls "the technium"?

Tucked away in a footnote on page 505 we find Kurzweil again equivocating: "The Turing test is intended as a measure of *human* intelligence; failure to pass the test does not imply a lack of intelligence" (Kurzweil 2005, 505 n.30). What is this other non-human intelligence? Artificial intelligence. But if we accept a real distinction here, then the computational capacity of AI isn't even in competition with genuinely *human* intelligence. It's a race among machines that can become unbelievably fast and capacious without ever passing the Turing test.

The risks we run in trying to become more like our machines are subtle, as Ellen Ullman shows us in her beautifully written and sensitive book, *Close to the Machine: Technophilia and Its Discontents* (1997). Just as Fredkin's wife, Joan, describes her husband as having "cultivated . . . an 'impersonal intelligence,' a detachment from day-to-day affairs that protected him from his own emotions" (Wright 1998, 20), so Ullman paints a portrait of life among the computer nerds in Silicon Valley that ought to give us pause:

> I'd like to think that computers are neutral, a tool like any other, a hammer that can build a house or smash a skull. But there is something in the system itself, in the formal logic of programs and data, that recreates the world in its own image. Like the rock-and-roll

> culture, it forms an irresistible horizontal country that obliterates the long, slow, old cultures of place and custom, law and social life. We think we are creating the system for our own purposes. We believe we are making it in our own image. We call the microprocessor the "brain"; we say the machine has "memory." But the computer is not really like us. It is a projection of a very slim part of ourselves: that portion devoted to logic, order, rule, and clarity. It is as if we took the game of chess and declared it the highest order of human existence. (Ullman 1997, 89)

But life is not like chess. It is much messier. We neaten it up with our technologies at our peril. Ullman, who dallied with Marxism in her youth, likens the dangers of techno-determinism to the rigors of the dialectic:

> During my days in the party, we used to say that Marxism-Leninism was a "science." And the party was its "machine." And when the world did not conform to our ideas of it – when we had to face the chaotic forces that made people believe something or want something or do something – we behaved just like programmers. We moved closer to the machine. Confronting the messiness of human life, we tried to simplify it. Encountering the dark corners of the mind, where all sorts of things lived in a jumble, we tightened the rules, controlled our behavior, watched what we said. We were supposed to want to be "cogs in a wheel." (Ullman 1997, 30)

Yes, we've been here before: megalomaniacal "programs" for the improvement of humankind. Didn't work so well for the Soviet Union, or Cuba. Perhaps we want to be a little cautious before trying to enhance humanity again.

This is not to say improvement is either impossible or wrong. In reviewing the literature of transhumanism, I find myself leaning away from the conservatism of Fukuyama (2003), the President's Council on Bioethics (2003), and Michael Sandel (2004), and more towards the work of Arthur Caplan (2004a, 2004b, 2009). Recall, I started by quoting Nietzsche, the philosopher of the *Übermensch.* But when it comes to choosing specific directions and techniques for enhancement, I find myself more drawn to the human potential movement than to the laboratories of computer scientists. Particularly in the work of Michael Murphy (Ogilvy 2010) there is a vision of possibilities for the human condition seen whole that reflects real wisdom rather than a narrow obsession with logico-analytic intelligence. True, there are parts of the human potential movement that fix on particular *sidhis* like telepathy or levitation that strike me as misguided. And there is a strong strain of denial of death in the preoccupation with the survival of bodily death. But on the whole, and for the sake of the whole, I think that the human potential movement has a lot to offer the transhumanist movement. Like Nietzsche, both movements see humanity as a bridge and not an end. We can do better. We can be better. But we must choose our tools and our techniques and our directions wisely.

References

Adorno, T.W., E. Frenkel-Brunswik, D. Levinson, R. N. Sanford. 1969. *The authoritarian personality.* New York: Norton. (Orig. pub. 1950.)

Anderson, P. 1972. More is different. *Science*, New Series 177, 4047: 393-96.

Bellah, R. with Richard Madsen, William Sullivan, Ann Swidler, and Steven Tipton, 1985. *Habits of the heart.* Berkeley, CA: University of California Press.

—. 1991. *The good society.* New York: Alfred A. Knopf.

Boring, E. G. 1923. Intelligence as the tests test it. *New Republic*, 6 June, pp. 35-37.

Brand, S. 1968. *The whole earth catalog.* Menlo Park: The Portola Institute.

Brooks, David. 2000. *Bobos in paradise: The new upper class and how they got there.* New York: Simon & Schuster.

—. 2011. *The social animal: The hidden sources of love, character, and achievement.* New York: Random House.

Caplan, A. 2004a. Nobody is perfect – but why not try to be better? *PLOS Medicine* 1:3, 52-54.

—. 2004b. Is biomedical research too dangerous to pursue? *Science* 303: 1142.

—. 2009. Good, better, or best. In *Human enhancement,* ed. J. Savulescu and N. Bostrom, 199-209. Oxford: Oxford University Press.

Deacon, T. 1997. *The symbolic species***.** New York: Norton.

—. Forthcoming. *Incomplete nature: How mind emerges from matter.* New York: Norton.

Dennett, D. 1984. *Elbow room: The varieties of free will worth wanting.* Cambridge, MA: MIT Press.

Diagnostic and statistical manual of mental disorders IV. 2004.

Dreyfus, H. 1992. *What computers still can't do.* Cambridge, MA: MIT Press.

Erikson, E. H. 1968. *Identity: Youth and crisis.* New York: W. W. Norton.

Fukuyama, F. 2003. *Our posthuman future: Consequences of the biotechnology revolution.* New York: Picador.

Gardner, H. 1983. *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

—. 1985. *The mind's new science: A history of the cognitive revolution.* New York: Basic Books.

—. 1999. *Intelligence reframed: Multiple intelligences for the 21st century.* New York: Basic Books.

Ghamari-Tabrizi, S. 2005. *The worlds of Herman Kahn.* Cambridge: Harvard University Press.

Herrnstein, R and C. Murray 1994. *The bell curve.* New York: Free Press.

Kahn, H. 1960. *On thermonuclear war.* Princeton: Princeton University Press.

Kelly, K. 2010. *What technology wants***.** New York: Viking

Kohut, H. 2000. *Analysis of the self: Systematic approach to treatment of narcissistic personality disorders*. Madison, CT: International Universities Press.

Kurzweil, Ray. 2005. *The Singularity is near: When humans transcend biology.* New York: Viking.

—. 1999. *The age of spiritual machines: When computers exceed human intelligence.* New York: Viking.

Laing, R. D. 1965. *The divided self.* Baltimore: Penguin.

Lanier, J. 2010. *You are not a gadget.* New York: Alfred A. Knopf.

Lasch, C. 1978. *The culture of narcissism.* New York: W. W. Norton.

Maslow, A. 1968. *Toward a psychology of being.* New York: John Wiley & Sons.

Minsky, M. 1985. *The society of mind.* New York: Simon & Schuster.

Mitchell, A. 1983. *The nine American lifestyles.* New York: Macmillan.

Murphy, M. 1992. *The future of the body.* Los Angeles: Jeremy Tarcher.

Nietzsche, F. 1954. *Thus spake Zarathustra*. In *The portable Nietzsche,* ed. W. Kaufmann, pp. 103-449. New York: Viking.

Ogilvy, J. 2010. *An actual man: Michael Murphy and the human potentials movement.* Berkeley: Minuteman Press.

President's Council on Bioethics. 2003. *Beyond therapy: Biotechnology and the pursuit of happiness.* New York: HarperCollins.

Reich, W. 1970. *The mass psychology of fascism,* tr. Higgins. New York: Farrar, Straus and Giroux. (Orig. pub. 1946.)

Riesman, D. with Nathan Glazer and Ruel Denney, 1950. *The lonely crowd.* New Haven: Yale University Press.

Riso, D., and R. R. Hudson. 2000. *Understanding the Enneagram.* New York: Houghton Mifflin Harcourt.

Rogers, C. 1989. *The Carl Rogers reader.* New York: Houghton Mifflin.

Sandel, M. 2004. The case against perfection. *The Atlantic Monthly,* April, 51-62.

Searle, J. R. 1980. Minds, brains, and programs. *Behavioral and Brain Science* 3, 417-57.

—. 1984. *Minds, brains, and science.* Cambridge, MA: Harvard University Press

—. 1992. *The rediscovery of the mind.* Cambridge, MA: MIT Press.

—. 1997. *The mystery of consciousness.* New York: New York Review Books.

—. 2002. *Consciousness and language.* Cambridge, UK: Cambridge University Press.

—. 2007. *Freedom and neurobiology.* New York: Columbia University Press

Slater, P. 1990. *The pursuit of loneliness.* Boston: Beacon Press.

Tarnas, R. 2006. *Cosmos and psyche.* New York: Viking.

Turkle, S. 1995. *Life on the screen.* New York: Simon & Schuster.

Ullman, E. 2005. *Close to the machine: Technophilia and its discontents.* San Francisco: City Lights Books.

Wright, R. 1998. *Three scientists and their gods.* New York: Times Books.

Zielenziger, M. 2006. *Shutting out the sun.* New York: Doubleday.