



PROJECT MUSE®

9. Artificial Intelligence and the Essence of Humanity

Published by

Goldberg, Steven.

Culture Clash: Law and Science in America.

NYU Press, 1994.

Project MUSE. <https://muse.jhu.edu/book/15796>.



➔ For additional information about this book

<https://muse.jhu.edu/book/15796>

Artificial Intelligence and the Essence of Humanity

Spectacular developments in the far reaches of computer science are announced almost daily. And the most striking advances concern not the brute power of ever faster machines, but artificial intelligence. Computers that appear to think will, we are told, revolutionize the workplace and the schoolroom. This field has so far produced relatively little in the way of usable technology, but it has triggered an enormous debate on philosophical questions about the nature of intelligence. As with so many debates in our society, most of the participation on both sides has taken the scientists' view of the world, even when implicit value questions cry out for the addition of other perspectives.

The Origins of Artificial Intelligence

Artificial intelligence, the discipline that seeks to build thinking computers,¹ traces its origins back to early efforts to have machines do mathematics. In the nineteenth century, Charles Babbage's "difference engine," built with financial support from the British government, was able to calculate the values of certain polynomials.² Babbage later conceived an "analytic engine" a general-purpose computing machine. Although

never completed, its design anticipated many features of modern computers.³

In Chapter 6 we traced the development of modern computers in this century. As the power of these machines grew

[136.0.111.243] Project MUSE (2025-01-18 20:26 GMT)

first-generation
digital computers
and their size

shrank, comparisons with the human brain became ubiquitous. The theoretical groundwork for thinking about the ultimate capability of computers was largely laid by Alan Turing, the brilliant British mathematician and codebreaker. Turing played a key role in developing the symbolic computation that is central to modern efforts at artificial intelligence. In 1936 he devised a hypothetical logic machine that consists of an endless piece of paper divided into squares, where each square is either blank or marked. A scanning device moves along the paper and either makes a mark, erases one, or moves one square forward or one back. Turing was able to show that this machine could compute anything that could be computed by any machine, no matter how complex.⁴ In other words, logic problems, and any problems that can be formulated logically, can be reduced to a series of yes or no steps.

Modern research in artificial intelligence was energized by a pivotal summer conference at Dartmouth College in 1956.⁵ John McCarthy, assistant professor of mathematics at Dartmouth, coined the term *artificial intelligence* while writing a proposal for the conference. At the gathering, McCarthy, Marvin Minsky, Allen Newell, Herbert Simon, and others exchanged findings and ideas about computer languages that were suited to flexible problem solving and about programs they had written that could, for example, prove mathematical theorems. To artificial intelligence researchers, computers are not number crunchers. They are machines that can mimic human behavior in areas ranging from conversation to chess.

Two major approaches characterize current work in artificial intelligence (AI). Symbolic AI relies on serial processing in which software operates on hardware in a single track approach to problems. Neural network AI, which is modeled more on how the human brain functions, uses parallel processing and does not distinguish as sharply between hardware and software.⁶ Both techniques pursue the ultimate goal of building machines, including robots with an array of sensory devices, that can learn and think in a human fashion.

Artificial Life

In recent years, a rival conception of computer intelligence has begun to challenge traditional AI. The artificial life movement is more concerned

with computer programs that mimic birds in flight than with those that play chess. Although often described as a subset of artificial intelligence,⁷ the history and goals of artificial life are somewhat distinct.

The intellectual forbearer of artificial life is John von Neumann, a mathematician who was born in Budapest and came to the United States in 1930. After World War II, von Neumann was one of the creators of the modern concept of the computer, in particular the linkage between a memory unit and a central processing unit. Von Neumann wrote quite generally about what he called *automata*, or self-operating entities, which could proceed "in light of instructions programmed within itself."⁸ Von Neumann included both machines and biological organisms in this category of "automata."⁹

This concept evolved in the late 1960s with the creation of computer games, such as *Life*, in which simple instructions when played out at length led to fascinating patterns of "cell birth" and "death" emerging on a sort of checkerboard on the computer screen.¹⁰ Influenced by these and other developments, Christopher Langton of the Los Alamos National Laboratory coined the term *artificial life* and gave the field definition by organizing the first artificial life conference at Los Alamos in 1987.¹¹

As presently practiced, artificial life involves writing programs that create flexible computer entities. These entities are given a few simple rules to follow that govern such matters as mobility and reproduction. When their environment changes they adapt. Thus computer-simulated "birds" have evolved "flocking" behavior that is remarkably similar to that of biological birds.¹² Artificial life proponents maintain that what they have created will someday deserve to be called "alive" every bit as much as creations made of DNA. As Langton has put it, "microelectronic technology and genetic engineering will soon give us the capability to create new life forms *in silico* as well as *in vitro*."¹³

At one level, the contrast with traditional AI is sharp. Artificial life seeks to replicate biological organisms from the bottom up through evolution, whereas AI operates from the top down by creating machines that undertake sophisticated activities like playing chess. Langton has written that the first conference on artificial life revealed that the participants shared "a very similar vision, strongly based on themes such as *bottom-up* rather than *top-down* modeling, *local* rather than *global* control, *simple* rather than *complex* specifications, *emergent* rather than

prespecified behavior, *population* rather than *individual* simulation, and so forth.”¹⁴

At another level, however, artificial life shares a great deal with artificial intelligence, and it is not surprising that at many universities they are grouped in the same department. The goal, after all, is to have computers, all of which operate within the limits set by Turing’s theoretical work, take on a variety of tasks previously thought to be the sole province of biological organisms.

The Practical Impact of Artificial Intelligence

The potential impact of artificial intelligence, broadly defined, ranges over fields like machine translation, medical diagnosis, modeling of diseases, and operation of aircraft. The U.S. government has been sufficiently intrigued to support AI with funding from a variety of sources. The Advanced Research Projects Agency within the Department of Defense has provided support, as have the individual services through the Office of Naval Research, the Army Research Office and the Air Force Office of Scientific Research. On the civilian side, the National Science Foundation funds and coordinates a variety of endeavors. The National Aeronautics and Space Administration, for example, is interested in the use of intelligent robots in space missions. And the National Institutes of Health has supported efforts to search for a cure for AIDS using artificial life techniques.¹⁵

Government support for AI has had its ups and downs but it will persist. Much of it is “little science,” spread out, in classic fashion, among a variety of agencies and in the civilian and military sectors. There is enough intellectual interest to engage the research community and enough promise of practical payoff to satisfy the legislature.

Of course, here, as elsewhere with basic science, the practical payoffs are more elusive than they first appear. AI has not been immune to the overpromising that so often occurs when the scientific ethic of progress is expressed in the public realm. In 1949, for example, the director of natural science at the Rockefeller Foundation proposed that computers be used to solve “worldwide translation problems,” reasoning that because they had broken codes in the war, they could “decode” one language into another.¹⁶ But language turned out to be a good bit more subtle and context dependent than supposed, and machine translators

have so far solved rather little. More recently, the U.S. Army's prize AI project, the "Autonomous Land Vehicle" ran off the road. Plans called for an armored transport vehicle that could deliver supplies under combat conditions without a human driver, but five years of work produced a vehicle unable to drive at adequate speeds even under test conditions, and the program was cancelled.¹⁷

At present, although many computer programs claim to be "intelligent" or to use "artificial intelligence," actual uses of computers for flexible problem solving typically done by humans have been rare in the commercial sector. Although many were launched with great fanfare, very few companies specializing in artificial intelligence have survived, and many of those are struggling.¹⁸ The main successes have been in narrow areas such as credit verification.¹⁹ Moreover, we can confidently predict that when products do arrive in large numbers, the regulatory gap will take its toll. If, for example, clever programs are written to diagnose diseases, that does not mean they were written with the sorts of questions that arise in malpractice actions in mind. And, as noted in chapter 6, the protection of intellectual property in computer software remains a formidable problem. Given the paucity of actual artificial intelligence products on-line, it is difficult to say exactly where other regulatory problems will come in, but there is no reason to believe they will be absent.

Consciousness and Human Uniqueness

All of this means simply that artificial intelligence is likely to follow the normal path to usable technology—rockier than anticipated. But as with other developments in science, the limited commercial impact does not mean limited impact on our values. Indeed, the prospect of thinking computers has created an extraordinary amount of public debate on remarkably philosophical issues. As computers have apparently become more like humans, we humans have struggled to retain our distinctiveness. This has resulted in a new emphasis on the importance of human consciousness and self-awareness, and a growing debate over whether machines can possess those traits. But, remarkably, even this debate is dominated, on both sides, by the scientific world view.²⁰

The observation that advances in science can diminish an individual's sense of uniqueness is hardly new. Sigmund Freud, for example, wrote

that “[h]umanity has in the course of time had to endure from the hands of science two great outrages upon its naïve self-love.”²¹ Freud identified these “outrages” as Copernican astronomy, which displaced the earth from the center of the universe, and Darwinian evolution, which posited a continuity between animals and man.²² Freud believed that resistance to his own theories stemmed from the fact that he further insulted humanity by “endeavouring to prove to the *ego* of each one of us that he is not even master in his own house, but that he must remain content with the veriest scraps of information about what is going on unconsciously in his own mind.”²³

If Freud correctly identified three blows to our pride, two of them have to some extent been absorbed. Man has ceded his place in the physical center of the universe with reasonable grace. And if Freud’s own theories have not achieved the acceptance of those of Copernicus, that may simply be because they do not deserve that acceptance on scientific grounds. In any event, many people today can comfortably concede that they sometimes have, as Freud argued, unconscious motivations.

Darwin’s theory of evolution remains much more controversial in some sectors of society, and that may be precisely because it continues to threaten human self-esteem in the way that Freud suggested. Although evolution is widely supported by the scientific community,²⁴ its inclusion in the curriculum remains a contentious subject in many schools, as we saw in chapter 5.²⁵ Too much continuity between mankind and other animals can be quite upsetting.²⁶

Even for many of those who believe in evolution, the human trait of self-awareness preserves a distinction between people and at least some animals. Consider, for example, the position of Peter Singer, the influential advocate for animal rights, whose *Animal Liberation* has been termed “the bible of the animal liberation movement.”²⁷ Singer presents a strictly utilitarian argument; following Bentham, he says the relevant question is not whether an animal can reason or talk, but whether an animal can suffer.²⁸ He thus stresses the abundant evidence that animals can feel pain and argues that this pain must be taken into account in human decision making.²⁹ But he expressly rejects the notion that humans are like all other animals in terms of self-awareness:

To avoid speciesism we must allow that beings who are similar in all relevant respects have a similar right to life—and mere membership in our own biologi-

cal species cannot be a morally relevant criterion for this right. Within these limits we could still hold, for instance, that it is worse to kill a normal adult human, with a capacity for self-awareness and the ability to plan for the future and have meaningful relations with others, than it is to kill a mouse, which presumably does not share all of these characteristics.³⁰

Indeed, Singer argues generally that nonhuman animals are not capable of making moral choices, thus making irrelevant the question of the propriety of their behavior toward each other.³¹ It is no criticism of animal rights advocates to note that some of them might find it more difficult to subscribe to a doctrine that did not retain such a special role for humankind. There are, after all, many in the emerging field of cognitive ethology who, on thoroughly Darwinian grounds, find evidence for animal consciousness and awareness and planning in animals ranging from birds to rhesus monkeys to honeybees.³²

In any event, whether or not humans feel comfortable being grouped with animals, a new challenge to human uniqueness has arisen with the development of the computer. As Bruce Mazlish put it, if Copernicus, Darwin, and Freud placed man on a continuous spectrum with the universe, the animal kingdom, and his own psyche, modern technology seeks to eradicate the final discontinuity—that between man and machines.³³

Preserving this discontinuity has become important to many people. As soon as machines master one task done by humans, there are those who say that some other task involves the really essential aspect of humanity. It is this progression that has led to the modern preoccupation with self-awareness.

The story of chess-playing computers, an important area of artificial intelligence research, is instructive. Although the rules of chess are easily taught to a computer, and although a computer can calculate much more rapidly than a human, it was long believed that certain essential intuitive aspects of the game would always give human players an advantage over a machine. As recently as 1987, one commentator claimed that "it has been calculated that a computer big and fast enough to beat a world class chess master would have to be approximately the size of the solar system."³⁴ By 1990 this claim had been falsified rather dramatically by the program Deep Thought, which has defeated several grandmasters and can run on any powerful computer.³⁵ Although Deep Thought has been defeated by the world champion, the program is

better than all but about one hundred players in the world.³⁶ Deep Thought plays much better than the five graduate students who developed it, and it will continue to improve.³⁷ According to one grandmaster who lost to it, "Deep Thought combines enormous speed and computational power with sophisticated analysis, itself developed by computer, of the relative values of the chess pieces depending on where they are and what stage the game has reached."³⁸

Yet some now explicitly reject the notion that grandmaster-level chess programs represent a milestone in the development of artificial intelligence.³⁹ To them, conscious self-awareness is a key aspect of the human brain's activity, and, they assert, such consciousness has not been achieved by current chess-playing machines.⁴⁰ In other words, Deep Thought may play a mean game of chess, but, unlike humans, it does not know it is playing chess.

Thus to many, the utterly internal, subjective sense of consciousness represents a safe harbor from the march of progress:

Science has revealed the secrets of many initially mysterious natural phenomena—magnetism, photosynthesis, digestion, even reproduction—but consciousness seems utterly unlike these. For one thing, particular cases of magnetism or photosynthesis or digestion are in principle equally accessible to any observer with the right apparatus, but any particular case of consciousness seems to have a favored or privileged observer, whose access to the phenomenon is entirely unlike, and better than, the access of any others—no matter what apparatus they may have.⁴¹

Thus the human capacity for self-awareness—the capacity to be conscious of our own existence—is still celebrated in modern culture. Even a fancy computer cannot think about itself the way we can ponder ourselves. Indeed, self-reference has been described as "America's latest social and pop-intellectual trend,"⁴² a trend exemplified by phenomena ranging from the "Gary Shandling Show" to Pulitzer Prizes for media coverage of the media.⁴³ Even law review articles have begun to appear that contain footnotes citing themselves.⁴⁴

The Debate over Computer Consciousness: Science versus Science

But to many proponents of artificial intelligence, consciousness is not a safe harbor inaccessible to scientific progress. They have argued explic-

itly that digital computers cannot be assumed incapable of any mental activity:

Minds exist in brains and may come to exist in programmed machines. If and when such machines come about, their causal powers will derive not from the substances they are made of, but from their design and the programs that run in them. And the way we will know they have those causal powers is by talking to them and listening carefully to what they have to say.⁴⁵

Thus the proponents of what has come to be called “strong” artificial intelligence believe that a properly programmed computer “would not only think but know it is thinking.”⁴⁶ The philosopher John Pollock, who is engaged in the OSCAR project to build a thinking machine, has written in *How to Build a Person* that “[t]here is no obstacle to building consciousness into an intelligent machine.”⁴⁷ Most dramatically, Hans Moravec, the director of the Mobile Robot Laboratory of Carnegie Mellon University, has maintained that he already sees evidence of awareness in his computer-driven mobile robots that use sensors to obtain information about their location and movements:

In our lab, the programs we have developed usually present such information from the robot’s world model in the form of pictures on a computer screen—a direct window into the robot’s mind. In these internal models of the world I see the beginnings of awareness in the minds of our machines—an awareness I believe will evolve into consciousness comparable with that of humans.⁴⁸

There has been a strong public reaction to claims such as these, and the growth of a remarkable and popular set of sophisticated arguments designed to show that digital computers are not conscious. By examining three of these opponents of strong artificial intelligence—the philosopher John Searle, the mathematician Roger Penrose, and the neurobiologist Gerald Edelman—we can witness the surprising role of science in value formation. Surprising, because the opponents of computer consciousness, no less than the proponents, stress the scientific world view.

John Searle’s analysis of the problem requires an understanding of the so-called Turing test. In 1950, when computers had very little power in modern-day terms, Alan Turing devised an operational test to determine when a computer could think like a human.⁴⁹ Under the Turing test, as this approach was inevitably named, a computer and a human are hidden from the view of a human interrogator.⁵⁰ The interrogator puts questions to the computer and to the human with a mechanism such as

a keyboard and screen. Any questions at all can be asked. The computer and the human answer through the keyboard and screen mechanism, but they have different goals. The computer has been designed to pretend that it is a human; the human is simply being him- or herself. At the end, interrogators have to decide which of the two entities they have been communicating with is the human. If they consistently get it wrong, that is, if the computer has consistently fooled them, we say the computer has passed the Turing test.

No computer can currently pass the Turing test, and a moment's reflection will persuade you that it will be quite a feat if a computer ever does. Obviously the interrogator will not rely on simple informational questions where a computer can be easily programmed to give a humanlike set of responses.⁵¹ (A set that might include, of course, incorrect answers to difficult math problems.)⁵² The skillful interrogator will rely instead on dialogues in which the computer will be forced to respond persuasively to questions about earlier parts of the discussion, to sarcasm, and to context-based comments on a variety of topics.⁵³ Nonetheless there are computers today that can engage in a reasonably sophisticated dialogue with humans, and they have fooled some people in Turing tests on limited topics.⁵⁴ Unsurprisingly, there are many in the artificial intelligence community who believe that a computer will someday pass a generalized Turing test.⁵⁵

The philosopher John Searle does not debate whether a computer could ever pass the Turing test. Instead he challenges the notion that it would be terribly meaningful if one could. Philosophers and others have long discussed the implications of having a machine that could pass the Turing test.⁵⁶ Searle's powerful contribution to the debate has made an unlikely appearance in popular books,⁵⁷ periodicals,⁵⁸ and even daily newspapers.⁵⁹ Indeed, Searle's "Chinese room" argument has become a flashpoint in the dispute over whether human consciousness is unique. With computers making human reason seem decidedly ordinary, there is a dramatic upsurge in interest in the question of whether computers can understand what they are doing in the sense that humans do. This interest is so strong that Searle, through no fault of his own, has had his views badly misrepresented at times. Indeed those who look to Searle for a vindication of human uniqueness are looking in the wrong place indeed. To understand why this is so, we have to begin in the "Chinese room."

Imagine, says Searle, that programmers have written a program that enables a computer to "understand" Chinese in the following sense.⁶⁰ If the computer is given a question written in Chinese characters, then it can produce an excellent answer written in Chinese characters. The computer is so good at this that it cannot be distinguished in a blind test from a native speaker of Chinese. This computer can be said to have passed a form of the Turing test, but is it correct to say that the computer understands Chinese in the way that the native speaker does? To answer this question Searle proposes the following thought experiment: Imagine you do not understand a word of Chinese. You are locked in a room with a large basket full of Chinese symbols written on bits of paper. Also in the room is a large book written in English that gives a series of rules for manipulating Chinese symbols. The rules say nothing about the meaning of the symbols; they simply say things like "if you are given a squiggle-squiggle symbol from outside the room, find a squoggle-squoggle symbol and pass it back out under the door." While you are in the room, someone outside starts passing Chinese symbols on bits of paper under the door. You follow the rule book and pass the appropriate bits with Chinese symbols back out under the door. As it happens, the bits coming in contain questions and the rule book you are following is cleverly designed so that you are sending out excellent answers. But you know none of this; moreover, you have no knowledge of what any of the symbols mean. Nonetheless, to an outside observer you will seem to have passed the Turing test. Just like the outstanding computer, your answers will be indistinguishable from those given by a native speaker. Yet, Searle concludes, you surely do not understand Chinese in the same sense as a native speaker; indeed, you do not understand Chinese at all. You are mindlessly manipulating symbols according to a book of rules. Similarly, Searle argues, the computer does not understand Chinese. It too is simply manipulating symbols according to a set of rules. Searle summarizes his point by saying that syntax alone is not sufficient for semantics.⁶¹

Searle's argument has created a firestorm of protest among the proponents of strong artificial intelligence.⁶² His argument is important in part because it is so general. It is not subject to persuasive criticism on the grounds that future advances in computer speed or in parallel processing will change the terms of the debate.⁶³ Searle has stressed that the essence of the strong artificial intelligence position is that binary processing

per se—breaking everything down into yes-no questions manageable by a digital computer—can give rise to consciousness.⁶⁴ Searle notes that, in principle, digital computers can be made out of anything at all—beer cans connected by strings, for example.⁶⁵ Proponents of strong artificial intelligence are committed to the position that a beer can computer of sufficient complexity could pass the Turing test and would be as conscious of what it is doing as a human.⁶⁶

The objections to the Chinese room argument are legion. Some contend that whereas the individual in the Chinese room does not understand Chinese, the whole system does.⁶⁷ Just as a single neuron does not understand anything, but the brain does, the individual does not understand Chinese, but the system—the individual, plus the rule book, plus the room, plus the slips coming in and out—understands Chinese.⁶⁸ Others argue that there is simply no way to know whether any other individual understands anything except by what they say or do, so we simply must assume that the person in the Chinese room understands Chinese just as we assume that sort of thing about each other.⁶⁹ Still others directly challenge the notion that syntax is not sufficient for semantics. They note that it was once believed that compression waves in the air could not produce sound and that oscillating electromagnetic forces could not produce light.⁷⁰ Both of these positions are now discredited.⁷¹ Someday we may realize, the argument goes, that a sufficiently complex syntactical system can indeed produce meaning.⁷² If you wave a bar magnet in a dark room you do not get any visible light, just as a person leafing slowly through what would be an extraordinarily large rule book does not appear to obtain any visible understanding.⁷³ In both cases the situation would appear quite different as our understanding of the processes involved increased.⁷⁴ And there are other responses to Searle as well.⁷⁵

This is not the place to evaluate the strength of Searle's argument. The crucial point from our perspective is that Searle has attracted popular attention because he seems to oppose those who would explain away human uniqueness. Thus the *New York Times*, after discussing the Chinese room, concluded that Searle questions "the premise of Western science: that the world we live in and the world inside our heads can be understood by the human mind."⁷⁶ Meanwhile, a law professor asserts that Searle has shown that "consciousness and intentionality . . . constitute human thought."⁷⁷ When computers start beating grandmasters at

chess, it is time to find a new discontinuity between man and machine. The subjective experience of human self-awareness has emerged as a likely candidate.

The irony here is that Searle himself, as he has continually emphasized, is hardly the man to defend the notion that there is some unknowable corner of the human mind or that humans somehow are fundamentally different from machines. Searle's argument applies only to the notion that binary computation alone can give rise to consciousness. He does not maintain that no machine can think, indeed, he has written:

[W]e are all machines. We can construe the stuff inside our heads as a meat machine. And of course, we can all think. So, in one sense of "machine," namely that sense in which a machine is just a physical system which is capable of performing certain kinds of operations, in that sense, we are all machines, and we can think.⁷⁸

Nor has Searle argued that only biologically based systems can think.⁷⁹ He has written explicitly that "[r]ight now [biologically based systems] are the only systems we know for a fact can think, but we might find other systems in the universe that can produce conscious thoughts, and we might even come to be able to create thinking systems artificially."⁸⁰ His point is that binary manipulation alone is not sufficient to create conscious thought. Indeed Searle, far from believing in some immaterial corner of the human mind, takes precisely the opposite position on the mind-body problem. An ardent opponent of dualism, he regards all mental phenomena, including consciousness, as being caused by processes going on in the brain.⁸¹ To Searle, brains cause consciousness in the same way that stomachs cause digestion.⁸² Searle accuses proponents of strong artificial intelligence as being victims of "a residual dualism."⁸³ No one would believe that a computer simulation of a stomach can actually digest anything, but people want to believe that a simulation of a brain can actually think. This is only possible, Searle asserts, if one believes the mind is independent of the physical brain, a classic dualist position.⁸⁴ We must, Searle concludes, "escape the clutches of two thousand years of dualism" and recognize that "consciousness is a biological phenomenon like any other."⁸⁵

Regardless of whether proponents of artificial intelligence are dualists or whether dualism is as bankrupt as Searle argues, the point is simply that the Chinese room offers no refuge for those who seek a special place

for humanity. Viewing humans as “meat machines” and consciousness as a “biological phenomenon” is hardly the approach for those who would retain a discontinuity between men and machines.

A similar irony is present in the public reaction to *The Emperor's New Mind* by noted mathematician Roger Penrose.⁸⁶ Penrose's work is a lengthy, complex treatment of a variety of topics in modern physics and mathematics. It is nonetheless a bestseller, a tribute in part to its underlying theme that digital computers are not the same as human minds.⁸⁷ As one commentator put it, people do not like “to see themselves as digital computers. . . . To be told by someone with impeccable scientific credentials that they are nothing of the kind can only be pleasing.”⁸⁸ One popular magazine puts the matter bluntly: “Those Computers Are Dummies” reads the headline, and the text asserts that “Penrose's central conclusion is that computers will never think.”⁸⁹

It is certainly correct that Penrose rejects the claim of strong artificial intelligence that digital computers can do and experience all that human brains can do and experience. But here, as with Searle, when we look more closely at his position we find little to comfort those who want a special status for humans.

Penrose criticizes strong artificial intelligence on different grounds than Searle does, although he starts in the same place. Penrose, like Searle, emphasizes that digital computers, no matter how complex, rely on step-by-step algorithms to solve problems.⁹⁰ Whereas that approach is adequate for an extraordinarily wide range of problems, Penrose maintains that the brain necessarily uses other approaches in certain of its activities.⁹¹ Unlike Searle, Penrose places particular stress on Gödel's undecidability theorem, which, he believes, demonstrates that humans can intuit as true certain propositions that cannot be established in a series of algorithmic steps.⁹² Finally, Penrose, unlike Searle, speculates that to fully understand how brains work and how consciousness arises, more will have to be learned about unresolved problems relating to quantum theory and other aspects of modern physics.⁹³

Thus it is clear, as Penrose explicitly states, that he does “not believe the *strong-AI* contention that the mere enaction of an algorithm would evoke consciousness,”⁹⁴ and for this conclusion Penrose has been criticized in the artificial intelligence community.⁹⁵ It has been argued, for example, that the neurons in the brain are simply too large to be affected by the quantum behavior Penrose stresses.⁹⁶ But whether Penrose is right

or wrong, the crucial point from our perspective is that he is not someone who maintains that there is something discontinuous between human beings and the rest of the physical universe. On the contrary, he seeks precisely the type of scientific unity typically sought by those who build grand theories. And in the end, he can see only two possibilities for understanding the conscious human mind. Either it grew out of "the thousands of millions of years of *actual* evolution that lie behind us,"⁹⁷ or it results from physical qualities that we may ultimately understand, thus enabling us "to construct such [conscious] objects for ourselves."⁹⁸ In the latter case—the human-built brain—the consequences could be dramatic:

One could imagine that these objects could have a tremendous advantage over us, since they could be designed *specifically* for the task at hand, namely to *achieve consciousness*. They would not have to grow from a single cell. They would not have to carry around the "baggage" of their ancestry (the old and "useless" parts of the brain or body that survive in ourselves only because of the "accidents" of our remote ancestry). One might imagine that, in view of these advantages, such objects could succeed in *actually* superseding human beings, where (in the opinions of such as myself) the algorithmic computers are doomed to subservience.⁹⁹

Thus the alternatives that Penrose sees are human minds as continuous with the animal kingdom through the mechanism of evolution or human minds as replicable, even surpassable, by human-made machines. Neither is a refuge for those who crave discontinuity.

Neurobiologist Gerald Edelman also challenges the notion of a thinking computer. His *Bright Air, Brilliant Fire* sets forth his view of the origins of human consciousness and of the debate over machine consciousness.¹⁰⁰ Like Searle and Penrose, Edelman rejects the idea that the human brain can be usefully likened to a digital computer. He argues, in a complex and controversial theory called neural Darwinism, that consciousness is an outgrowth of human evolution working at the level of groups of neurons reacting to sensory inputs. Edelman distinguishes between what he calls primary consciousness—an awareness of immediate surroundings without a sense of past or future—from higher order consciousness, which involves a sense of self and of time. The former may exist in some animals, such as dogs, whereas the latter is probably limited, in Edelman's view, to humans.

Reactions to Edelman's book have emphasized the point that Edel-

man rejects the strong claims made by some in the artificial intelligence movement. Edelman, we are told by journalists, argues that "the brain is not like a computer,"¹⁰¹ that the "richness of human experience could not be fitted into a mechanical or computer theory of the nervous system."¹⁰²

But, as with Searle and Penrose, Edelman is not, in fact, an unambiguous supporter of human uniqueness. He does emphasize that we come from a particular evolutionary background and are, in that sense, unlike other creatures. But, as an opponent of dualism, he sees the mind as a subject for scientific study in the ordinary sense. He believes that our higher order consciousness may have its origins in chimpanzees who have not just concepts, but some elements of self-concept.¹⁰³ Most strikingly, like Searle and Penrose, he sees no reason in principle why artifacts could not be constructed with high-order consciousness.¹⁰⁴ He thinks the day is far off, but concludes that "[m]y personal belief is that the construction of conscious artifacts will take place."¹⁰⁵

Our discourse is so dominated by scientists that our hunger for an alternative to humans as machines is fed by antidualist philosophers, mathematicians, and evolutionary biologists. There are in fact still thoughtful people, religious and otherwise, who believe the mind is in a fundamental way different than the body, but their voices are not loud.¹⁰⁶ There is an echo of those lost voices in John Updike's *Rabbit at Rest*. When Rabbit Angstrom resists heart surgery, his friend asks "What's wrong with running your blood through a machine? What else you think you are, champ?" Rabbit responds to himself, "A God-made one-of-a-kind with an immortal soul breathed in. A vehicle of grace. A battlefield of good and evil."¹⁰⁷

The scientific efforts to explain the mind should go on and they should be a vital part of public discourse. But there is room as well for discussions of the soul, and for nonscientific visions of the essence of humanity.

Consciousness and the Legal Definition of Death

The question of the uniqueness of human consciousness is of particular importance because it may play a role in one of the most vexing legal and social controversies of our time—the cessation of medical treatment and the definition of death. As Robert Veatch notes, any concept of the

death of a person depends directly on those qualities thought to make humans unique.¹⁰⁸ By that standard the modern trend is clear—in the space of a few decades technology has pushed us from a world in which a beating heart symbolizes life to a world in which heartbeat, breathing, eating, and even responding to external stimuli are less important than human consciousness. The question raised by developments in computer science is what might happen if we no longer viewed consciousness as unique to humans.

In the first half of this century, the interdependence of breathing, blood circulation, and the brain made the determination of death relatively uncontroversial.¹⁰⁹ The absence of breath and a heartbeat signified death.¹¹⁰ Beginning in the 1950s, however, artificial respirators and other life-support systems began to change the situation.¹¹¹ It became possible to keep the body alive when the brain had ceased functioning.¹¹² Indeed, it gradually became possible to replace virtually every part of the body except the brain with an artificial substitute.¹¹³

It may be difficult for us to recall, but these developments caused a tremor in mankind's sense of self. A government commission discussed "the problem of the 'man-machine symbiosis'—that is, the extent to which technological processes should be imposed upon, or substituted for, the natural processes of human beings."¹¹⁴ The focus of controversy was on the artificial heart. As the government commission put it, "The heart has held pre-eminence in poetry and in common speech as the seat of bravery, love, joy, and generosity. Will its replacement by a mechanical pump and motor not merely place technology deep in man's bosom but place man more deeply in the bosom of technology?"¹¹⁵

But the development of the artificial heart and other mechanical life-support devices hardly forced mankind to admit equivalence with machines. The human thirst for uniqueness was easily satisfied by moving all important characteristics to the one irreplaceable organ, the brain. A focus on the heart was dismissed as "symbolism" and as "irrational."¹¹⁶ As one ethicist put it in response to concerns over the artificial heart, "One can understand and psychologically become adjusted to the fact that the heart is a vital organic pump, and that it is not the inner core of the 'self.'"¹¹⁷

But the transfer of concern to the brain left difficult problems for medicine and law. When society was first confronted with comatose individuals whose breathing and heartbeat were sustained by machines,

it was not clear how to proceed; on the one hand, no one wanted to end the existence of someone who was likely to regain brain functioning, but on the other hand, the sense of respect for human dignity, the cost of medical treatment, and the desire for organs that might be transplanted into others all counseled for ending life support in certain cases.¹¹⁸ The result was a growing interest in "brain death," a concept that permitted death to be declared while the heart was still beating.¹¹⁹

In this country, the first major step toward defining death in terms of the brain was the report of an ad hoc committee at the Harvard Medical School in 1968.¹²⁰ The committee emphasized that with modern technology, respiration and a heart beat could be maintained "even when there is not the remotest possibility of an individual recovering consciousness following massive brain damage."¹²¹ Accordingly, the committee proposed standards for determining when patients should be declared dead because their brain was permanently nonfunctioning.¹²²

It immediately became apparent, however, that the word *brain* was arguably too broad for the purpose of defining death. The Harvard committee itself recognized the distinction between cortical and brain stem functioning.¹²³ Generally speaking, consciousness and cognition are carried on in the higher brain, that is, in the cerebrum, particularly the neocortex.¹²⁴ By contrast, vegetative functions such as breathing and blood pressure are carried out by the brain stem, a portion of the lower brain.¹²⁵ The distinction matters because in many heart attacks and accidents the disturbance of circulatory or respiratory functions is too brief to destroy the brain stem but sufficient to destroy the neocortex.¹²⁶ The result is an individual who is alive under a "whole brain" definition of death, because the lower part of the brain still works, but who would not be alive if a "higher brain" definition were used.¹²⁷ Such individuals are often in what is termed a persistent vegetative state.¹²⁸ They can be kept alive with intravenous feeding and antibiotics, but they are unlikely ever to recover further.¹²⁹ These individuals need no mechanical aid to breathe, maintain a heartbeat, react to light, or engage in other automatic functions, but they lack all awareness and thought.¹³⁰

As early as 1971, an editorial in *Lancet* referred to studies of comatose individuals who breathe but have no cortical functioning.¹³¹ The editorial concluded that because "death of a human being means death of his [her] brain—indeed of his [her] mind" an individual would not

want his or her vegetative existence to be prolonged after his or her cortex was destroyed.¹³²

An understanding of the legal regime that grew out of these developments requires that you keep in mind how rapidly things were changing. In the space of a few decades, the centuries-old identification of death with the cessation of the heartbeat and of breathing was giving way to a brain-centered definition. Fundamental moral issues were being addressed at the same time that new diagnostic techniques were being developed. As we shall see, the net result was that the legal definition of death moved to recognition that death of the whole brain meant death of the human being, even if the heart and lungs were still being artificially maintained.¹³³ At the same time, decisions on the cessation of treatment, including the cessation of mechanical feeding,¹³⁴ recognized that treatment could end not only when the whole brain was dead, but also when only the higher brain was destroyed.¹³⁵ It was, after all, the higher brain's ability to sustain consciousness that marked the uniquely human trait.

The legal definition of death moved quickly to reflect the considerations set forth in the report of the ad hoc Harvard committee. An influential step in this direction was the 1972 article by Capron and Kass proposing a model brain death statute.¹³⁶ Drawing in part on the Harvard committee's report, the authors proposed that if artificial means were keeping respiration and circulation going, a person would be considered dead if "he [she] has experienced an irreversible cessation of spontaneous brain functions."¹³⁷ The authors made clear that by "brain functions" they meant the whole brain, so that someone who had lost only higher brain functions would not be dead.¹³⁸ But the authors did not defend the proposition that lower brain functioning alone—a condition marked by spontaneous reflexes but no consciousness—constituted human life: on the contrary, they admitted that "the exclusion of patients without neocortical function from the category of death may appear somewhat arbitrary."¹³⁹ They defended the exclusion on the ground that they were taking a "modest" step to bring the definition of death in line with modern medicine, and they emphasized that modern medicine was not yet able routinely to diagnose irreversible higher brain death as clearly as whole brain death.¹⁴⁰ Moreover, they left the door open to the notion that individuals with higher brain death should

be allowed to die by stressing that they were discussing the question "is he dead?" not the question "should he be allowed to die?"¹⁴¹

Two influential reports by the Presidents' Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research took the same tack. The first, a 1981 report titled (in short) *Defining Death*, opted for a whole brain rather than a higher brain definition of death.¹⁴² The second, a 1983 report titled *Deciding to Forego Life-Sustaining Treatment*, said that families could justifiably remove artificial feeding tubes from patients with no higher brain functions in order to cause the death of those patients.¹⁴³

The law has generally followed these approaches. In virtually every jurisdiction, statute or common law provide that individuals are dead if their whole brain has ceased functioning, even if breathing and circulation are artificially maintained.¹⁴⁴ At the same time, prior to the *Cruzan* decision, "an unbroken stream of cases has authorized procedures for the cessation of treatment of patients in persistent vegetative states," that is, patients who have suffered higher brain death.¹⁴⁵ Many of these cases involve the removal of feeding tubes.¹⁴⁶ The courts have used various theories, ranging from an assessment of the patient's previously expressed wishes to an analysis of the benefits of continued treatment,¹⁴⁷ but the results have been the same. Even outside the realm of litigation, many physicians, despite the long standing medical presumption in favor of treatment,¹⁴⁸ have come to believe that withdrawal of food and water from permanently unconscious patients is appropriate.¹⁴⁹

So in just the space of a few decades, a remarkable consensus has grown up that those who have permanently been deprived of self-awareness by cessation of higher brain functioning can be allowed to die. And the emphasis on the higher brain has clearly been driven by a concern for those qualities that make humans special. The cases upholding, on various theories, the withdrawal of diverse treatments from those in a persistent vegetative state are replete with references to this concern. In the first widely reported case of this type, that involving Karen Ann Quinlan, the court stressed medical testimony that Ms. Quinlan was "totally unaware," and she lacked those brain functions that are "uniquely human."¹⁵⁰ In its seminal 1986 *Brophy* decision, the Supreme Judicial Court of Massachusetts, in allowing removal of a tube providing food and water from an individual in a persistent vegetative state, argued that in such cases "the burden of maintaining the corporeal existence

degrades the very humanity it was meant to serve.”¹⁵¹ In 1988, the court in *Gray v. Romeo*, presented with the same facts as *Brophy*, reached the same conclusion: “The facts in [this] case support the finding that [the patient] would consider that efforts now to sustain her life demean her humanity.”¹⁵²

At the same time, a wide variety of ethicists have also concluded that the absence of higher brain functions deprives individuals of their essential humanity.¹⁵³ Many of these writers would not only allow the withdrawal of treatment from such individuals, they would declare the individuals dead.¹⁵⁴ There is a remarkable agreement in these writings on the essential importance of human self-awareness. Moreover, this view grew up very quickly after modern developments in medical technology made it possible to support the body when the brain was dead. Recall that the ad hoc Harvard committee issued its initial cautious report on determining brain death in 1968. Veatch, himself a supporter of the higher brain definition of death,¹⁵⁵ reports that in 1970 when Henry Beecher was asked what was essential to the definition of human life, he replied simply, “Consciousness.”¹⁵⁶ Since that time several philosophers discussing death have identified the essential human characteristic as the consciousness and self-awareness that resides in the higher brain. In 1975, William Charron argued for a “psychological definition of death” that focused on “irreversible loss of all consciousness.”¹⁵⁷ Five years later, Michael Green and Daniel Wikler contended that because personal identity does not survive death of the conscious brain, such death constitutes death of the person.¹⁵⁸ In 1981, Allen Buchanan, relying on a “cognitivist” concept of life that turns in large measure on “self-awareness,” concluded that a higher brain definition of death was “inescapable.”¹⁵⁹ Finally, in 1986, a law professor, David Randolph Smith, argued that the legal definition of death should be “neocortical death” because of the “centrality of consciousness and cognition as the quintessential attributes of human life.”¹⁶⁰

Although Smith may have lost the legal battle, his views have won the war. As we have seen, an individual in a persistent vegetative state is not legally dead (because the whole brain is not dead), but treatment, including artificially introduced food and water, will almost inevitably be removed from that individual until legal death takes place.¹⁶¹ Indeed, an outsider to our society would be justified in concluding that in practice, if not in our codes, we treat those with an irreversible loss of conscious-

ness as dead. We do not immediately bury them, but immediate burial is not our practice in many cases. When a hospitalized individual on a respirator suffers a fatal heart attack we do not bury that person with the respirator running—we remove the respirator, we wait for relatives, we make arrangements.¹⁶² In our culture there are moral constraints on our conduct toward those who are no longer living.¹⁶³ An outsider might conclude that when an individual is dead because of a permanent loss of consciousness, our customs require a solemn determination that the loss has taken place, a removal of feeding tubes, and then a burial.

The U.S. Supreme Court's much-publicized decision in *Cruzan v. Director, Missouri Department of Health*¹⁶⁴ did not end this trend. The Court, in *Cruzan*, allowed the state of Missouri to insist there be clear and convincing evidence of an individual's intent before food and water is terminated when someone is in a persistent vegetative state. This was seen by some as a setback for the "right to die" because it prevented for a time the termination of treatment for Nancy Cruzan. But only for a time. After the Supreme Court's decision, Cruzan's parents, presenting new evidence that Nancy would not have wanted treatment in her current state, petitioned the Missouri courts to remove her feeding tube on the ground that the evidence of her intent was now clear. Their petition was granted, the tube was removed, and Nancy Cruzan died.¹⁶⁵ Thus even under a demanding legal standard, our culture continually opts for the centrality of consciousness.

But what if a consensus developed in our society that computers are indeed self-aware in the same sense as humans? It is unlikely to happen in the foreseeable future, but such a consensus might arise someday concerning digital computers or it might accompany the development of other machines, such as those discussed by Searle, Penrose, or Edelman. Perhaps if that day ever comes we would finally cease trying to find distinctions between ourselves and our machines.¹⁶⁶

Western history suggests, however, that we would not give in so easily.¹⁶⁷ More likely, we will react by seeking some new trait we do not share with conscious computers. If the artificial heart did not phase us for long, perhaps the artificial mind will not either. One candidate for a uniquely human characteristic beyond self-awareness is already available. Some ethicists have described the capacity for social interaction as an essential human trait that is very different from mere conscious-

ness.¹⁶⁸ This capacity is central to one observer's view of why playing chess with the grandmaster program Deep Thought is not the same as playing with a human:

My father taught me chess when I was six or seven years old. We played often during winters when little work could be done at the farm. Comfortably encamped in front of the living-room stove, a large bowl of freshly popped corn nearby, we engaged in mortal combat as snow fell outside and cold winds howled.

I just cannot envision so relaxing and enjoyable a scene were Deep Thought's monitor substituted for my father.

More recently, on the evening I learned my grandparents were seriously ill, a friend invited me to play chess so I would not have to be alone just then, but also wouldn't be compelled to talk much if I didn't feel like it.

"Deep Thought" is incapable of a simple act of friendship and kindness such as that. Nor could the machine have given me the emotional support I needed at that difficult time.

Give me . . . mere mortals . . . with whom to play chess—for the simple pleasures by which they enrich my life—win, lose, or stalemate. Surely the machine has not been nor ever can be invented to improve upon that.¹⁶⁹

Thus computers force us to see chess playing, once regarded as a remarkable intellectual feat focusing on introspection and intuition, as really an occasion for human interaction. Even if Deep Thought knows it is playing chess, it is not good at discussing how its human opponent feels about the game. At least until computers are built that can provide solace for humans, this step would give us some psychological distance from machines. But if we know anything from recent history, moving the essence of humanity in this direction could have profound implications for the legal definition of death. If computers push our sense of uniqueness away from self-awareness toward social interaction, the definition of death may eventually follow. The groundwork has already been laid—without reference to machine self-awareness, law review articles have already proposed that human interaction should supplant mere consciousness as the legal standard for life. Kevin Quinn suggests that the definition of life should be set at a "threshold . . . higher than minimal consciousness": the standard supported is "a minimum capac-

ity for interpersonal relationships."¹⁷⁰ When that capacity is irretrievably lost, life-sustaining treatment may be removed.¹⁷¹ Nancy Rhoden contends that when "analysis of the patient's capacities shows that she [he] is unable to enjoy the distinctly human pleasures of relating to her [his] environment and to others, then family discretion is warranted [on whether to end treatment]."¹⁷² Rhoden states explicitly that such patients have capacities beyond those of an individual in a persistent vegetative state.¹⁷³

The adoption of these views is much more likely to come about if a consensus grows that computers have self-awareness. Of course, such a consensus, if it developed at all, would grow slowly and would never be complete. Absolute proof that another human is self-aware, let alone a machine, is not available; consciousness is intrinsically a subjective phenomenon. But we treat other humans as though they were conscious and we may come to act that way toward certain machines. It would be a gradual process as we came to think of particular devices as not merely doing things but as knowing they were doing things. As we have seen, nothing that Searle, Penrose, or Edelman proposes suggests that machines generally must lack self-awareness. If consciousness is merely a product of a machine—indeed, if, as Searle himself says, consciousness is to brains as digestion is to stomachs¹⁷⁴—consciousness will come to be viewed as no more special for the definition of life than breathing or circulating blood.

A diminished sense of the uniqueness of consciousness and a corresponding broadening of the category of those regarded as dead would also happen gradually, as change in the law generally does. Most likely, we would begin by continuing to call those who lack the capacity for social interaction alive, while we remove their means of existence, as we have done for those who lack consciousness.¹⁷⁵ But however gradually, the trend could move in this direction. There are after all, other pressures pushing toward the same end; many have long predicted or feared, for example, that social worth and economic considerations could force withdrawal of treatment from ill geriatric patients.¹⁷⁶ We now must recognize that the emergence of an increasingly rarified sense of what it is to be human could have the same impact. Indeed, this new pressure could prove to be the most important. One can argue rationally about the social justice issues involved in deciding which people are entitled to expensive treatment. But when one begins to view the subjects of treat-

ment not as people, but as indistinguishable from machines, the outcome of the debate is likely to be foreordained.

A view of what this future would look like is provided by Quinn and Rhoden's analysis of the *Conroy* case.¹⁷⁷ In *Conroy*, the Supreme Court of New Jersey, the court that decided the *Quinlan* case, considered whether a feeding tube could be removed from an elderly brain-damaged woman who was hospitalized with a variety of serious permanent ailments and who had a limited life expectancy.¹⁷⁸ Ms. Conroy could not speak or interact with other people, but she appeared to retain consciousness—she would, for example, scratch herself, pull at her bandages, and smile when her hair was combed or when she received a “comforting rub.”¹⁷⁹ The court explicitly noted that she was not in a persistent vegetative state.¹⁸⁰ The court concluded that the feeding tube could not be removed.¹⁸¹ Ms. Conroy had not made clear that she would have wanted such a result, nor was the court satisfied that the burdens of her life outweighed the benefits or that the pain of her life rendered treatment inhumane.¹⁸²

Quinn challenges this result on the ground that because an individual in Ms. Conroy's situation cannot engage in “interpersonal relationships” she has “no possibility of personal life” and thus there is no obligation to maintain “biological life.”¹⁸³ Rhoden explains her support for removing Ms. Conroy's feeding tube in somewhat different terms.¹⁸⁴ Ms. Conroy, she suggests, is not “able to experience or enjoy life.”¹⁸⁵ The court in *Conroy*, when it said feeding could not be terminated, “in essence reduced Ms. Conroy, or the person that she was, to an object that passively experienced physical sensations.”¹⁸⁶ It is easy to see how this view would gain strength from the development of machines we regard as self-aware. Viewing Ms. Conroy, a conscious individual, as “an object” is a tendency that can only be strengthened when we have “objects,” such as thinking machines, that we view as conscious.

Scientific developments, as we have seen, do not always lead in a direct and obvious way to practical applications.¹⁸⁷ When Einstein first developed the special theory of relativity he had no idea it would lead to nuclear power.¹⁸⁸ The impact of science on law is even more complex and uncertain. But it is no less real. Developments in artificial intelligence could change not only the machines we use, but our views of ourselves. And when those views change, our views of the proper legal standards for life and death could change as well.

The impact of artificial intelligence on legal doctrine is not as obvious as the impact of a medical innovation, such as the heart-lung machine. But it is not just medical devices that produce legal change. Darwin's theory of evolution had a dramatic impact on the development of the animal rights movement, a movement that seeks changes in the laws governing the treatment of animals.¹⁸⁹ In the early part of this century psychoanalysis shaped developments in the insanity defense and in the criminal law generally.¹⁹⁰ Given this history, artificial intelligence—the latest field to challenge humanity's sense of uniqueness—clearly has the potential to shape the legal standard for death.

There are, as we have noted, other factors pushing toward an enlarged definition of death. The enormous costs of medical treatment for the elderly and sick may, for example, move society in that direction.¹⁹¹ In an issue arising at the beginning of life, it has been argued that anencephalic infants—those born with no higher brain, but with rudimentary brainstem activity—should be declared dead so that organ donation will become possible.¹⁹² But multiple causation is hardly unusual in the law. The legal changes involving animal rights and those involving the insanity defense were not solely the result of Darwin and Freud.¹⁹³

It is speculative at best to suppose that we will, in fact, see developments in artificial intelligence that will alter our view that human consciousness is unique, and even more speculation is necessary to suggest that these developments will alter our legal standard for human life. But this kind of speculation is useful to show where our public debate is going. At present our discussions of human consciousness are dominated on one side by artificial intelligence experts who see the possibility of consciousness in digital computers, and on the other by philosophers and scientists who believe different artifacts would be needed to achieve consciousness. Religious perspectives link human consciousness to a nonorganic mind and human uniqueness to an immortal soul,¹⁹⁴ but those perspectives are absent in this discourse. Moreover, if public attitudes change and consciousness is seen as nothing special, it will not be the fault of the artificial intelligence movement or of its secular opponents (such as Searle, Penrose, and Edelman) if this eventually leads to an expanded definition of death. Their arguments are focused on the scientific capabilities of machines. They are not discussing legal and moral human obligations. It is not their fault if public silence elsewhere

allows science to carry weight far from its appropriate jurisdiction. Let this be an early warning—important values are at stake. The definition of death and the sanctity of human life should not turn on whether a digital computer or any other device appears to be conscious. But, farfetched as it may seem, that is a risk we currently run.