

MINDS AND MACHINES: A MODEST PROPOSAL

Albert A. Anderson

Professor of Philosophy Emeritus

Babson College

The first version of this paper was written in 1968. It was published in *Philosophic Research and Analysis* (Volume VII, No. 8, Spring, 1979). The present paper has been edited for stylistic purposes, but the content remains essentially the same as in the 1968 version.

MINDS AND MACHINES: A MODEST PROPOSAL

1. The Problem of Human Uniqueness

The purpose of this paper is to take a fresh look at the problem of where human beings fit into the natural order and to venture a hypothesis about the future direction of evolution.

In seeking to understand the place of human beings in nature, philosophers have sought to establish human uniqueness by showing the difference between nature and animal nature. In the 17th century, Descartes regarded animals as *automata*, self-moving machines that differ in kind from humans who possess both mind and soul. In the 19th century, evolutionary theory established obvious connections between human beings and their pre-human ancestors. Most contemporary thinkers exercise caution when suggesting such a dichotomy between human beings and their biological relatives.

Even Teilhard de Chardin, a Roman Catholic priest who was also a paleontologist, emphasized the unity of nature at every stage, showing the origin and development of life and mind as part of a monistic, harmonious process in which nature does not make any leaps. Teilhard's central thesis in *The Phenomenon of Man* is that the evolution of life displays a progressive development in which any advancement in the level of consciousness of an organism is accompanied by a more complex and better-organized nervous system.¹ It is impossible to say at what point mind (what Teilhard calls the noosphere) arises. The distinction between human beings and their immediate predecessors is far from the clean cut postulated by Descartes.

Nevertheless, the custom is still widespread to follow Aristotle's lead and say that humans are *rational* animals, thus separating them from other animals. The outward and visible sign of this rationality is the ability to use symbolic language. Descartes took his stand on this difference. If the line between human and pre-human life forms has become blurred, the separation between speaking and non-speaking animals is still a distinction that many find convincing. Even if other animals manifest pre-rational activity, the difference between linguistic and non-linguistic creatures establishes a comfortable distance between human beings and other animals.

Human dignity and human uniqueness are preserved even in light of recent experiments with chimpanzees that use sign language and dolphins that seem to communicate what they know. The culture-bearing human animal not only knows, but it knows that it knows. Teilhard argues that if other animals had that power, we would surely have observed it.² What most clearly separates human beings from earlier stages of evolution is that humans are conscious of the evolutionary process itself.

The twentieth century brought a new difficulty. Even if we clearly distinguish ourselves from other animals, the new question centers on how we differ from machines. Let it be said from the outset that this is not a question about present computers and so-called "thinking machines." I am well aware of the shortcomings of present mechanisms — even the most advanced ones. I am asking a philosophical question, not a technological

one. The question is whether there are any differences between human beings and machines that make it impossible *in principle* for machines to equal or even surpass human beings in intellectual achievements. On what grounds, if any, is it possible to distinguish not only present machines but also all future machines from human beings?

Let's turn again to Descartes. In 1637, when he wrote *Discourse on the Method*, Descartes offered two "certain" methods of distinguishing humans from machines.

I made special efforts to show that if any such machines had the organs and outward shape of a monkey or of some other animal that lacks reason, we should have no means of knowing that they did not possess entirely the same nature as these animals; whereas if any such machines bore a resemblance to our bodies and imitated our actions as closely as possible for all practical purposes, we should still have two very certain means of recognizing that they were not real men. The first is that they could never use words, or put together other signs, as we do in order to declare our thoughts to others. For we can certainly conceive of a machine so constructed that it utters words, and even utters words which correspond to bodily actions causing a change in its organs (e.g. if you touch it in one spot it asks what you want of it, if you touch it in another it cries out that you are hurting it, and so on). But it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do. Secondly, even though such machines might do some things as well as we do them, or perhaps even better, they would inevitably fail in others, which would reveal that they were acting not through understanding but only from the disposition of their organs. For whereas reason is a universal instrument which can be used in all kinds of situations, these organs need some particular disposition for each particular action; hence it is for all practical purposes impossible for a machine to have enough different organs to make it act in all the contingencies of life in the way in which our reason makes us act.³

These are the same grounds on which Descartes distinguished people from animals, which should come as no surprise when we recall that he regarded animals as "machines" (automata).

2. Thought and Communication

Contemporary philosophers have given much attention to the relationship between language and thought, and there is considerable disagreement over the matter. Descartes' assumption that "words" or "other signs" are necessary for communicating thoughts deserves analysis, but what I wish to consider here is whether his claim that machines could not "produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence." Today's machines seem to do just that.

Norbert Wiener, the great pioneer in the science of cybernetics, made great strides in clarifying the process of communication (a) between people and machines, (b) between

machines and people, and (c) among machines. The term “cybernetics” is derived from the Greek word *kubernan*, which means “to govern.” The act of governing, guiding, and steering is carried out through the process of communicating information. What activity is more indicative of the thinking process than that of ruling or guiding?⁴

According to Wiener, thinking depends on the ability to discriminate and manipulate patterns. We can consider the world itself to be made of patterns. Wiener puts the matter this way:

A pattern is essentially an arrangement. It is characterized by the order of the elements of which it is made, rather than by the intrinsic nature of these elements. Two patterns are identical if their relational structure can be put into one-to-one correspondence, so that to each term of the one there corresponds a term of the other . . . The simplest case of one-to-one correspondence is given by the ordinary process of counting.⁵

As Wiener analyzes it, what is most important for understanding communication is that the patterns of language are isomorphic with the patterns of the world. Linguistic patterns convey information, something transmissible from individual to individual. The pattern may be spread out in space, or it may be distributed in time. The pattern of a piece of wallpaper, for example, is extended in space, whereas the pattern of a musical composition is spread out in time. Wiener compares the pattern of a musical composition to the pattern of a telephone conversation and the dots and dashes in a telegram.⁶ These two kinds of pattern are designated as messages, not because the pattern of the conversation differs from the musical composition, but because it is used in a different way, “to convey information from one point to another.”⁷

Although there are more subtle and complex relationships in more advanced forms of communication, Wiener claims that the transmission of information by a machine and by a human being is the same *in principle*. This leads to his central claim:

It is my thesis that the operation of the living individual and the operation of some of the newer communication machines are precisely parallel. Both of them have sensory receptors as one stage in their cycle of operation: that is, in both of them there exists a special apparatus for collecting information from the outer world at low energy levels, and for making it available in the operation of the individual or of the machine. In both cases these are not taken *neat*, but thorough the internal transforming powers of the apparatus, whether it be alive or dead. This information is then turned into a new form available for the further stages of performance. In both the animal and the machine, this performance is made to be effective on the outer world. In both of them, their *performed* action on the outer world, and not merely their *intended* action, is reported back to the central regulatory apparatus.⁸

The importance of Wiener’s work in cybernetics is that he opened the way for contemporary machines to use language. They manipulate signs in ways that go far beyond the push-button stimulus and response behavior cited by Descartes concerning the machines of his day.

Consider this example:

In a program written by Gelernter, a computer can be set to seek the proof of a theorem in geometry, the same sort of problem that might give a bright high school student considerable food for thought and cause a less gifted one to give up entirely. The computer . . . will begin by trying some simple rules of thumb. Should these fail, the computer will formulate some conjecture that would advance the solution if it could be proven true. Having made such a conjecture, the computer will check its plausibility in terms of an internal diagram of the situation. If the conjecture is plausible, its proof is sought by the same rules of thumb as before. Once proved, the conjecture will serve as a steppingstone to the desired theorem. If the conjecture is rejected as implausible . . . others will be tried until one has succeeded or the computer's resources are exhausted. Not even the programmer knows in advance whether the machine will succeed in proving the given theorem. The number of steps involved is so great that their endpoint cannot be predicted. I would not deny that the computer has behaved intelligently. Avoiding blind trial and error, it has selected and pursued promising hypotheses.⁹

Based on this example alone, it we can conclude that computers and programs already exist that clearly and distinctly demonstrate that machines can manipulate abstract signs in ways that rival moderately intelligently people.

This is but a stumbling beginning compared with the possibilities being suggested. Hiller and Isaacson, in an article entitled "Experimental Music,"¹⁰ report the results of their computerized study of musical structures in terms of information theory. Their goal is to understand the basis of musical composition from the standpoint of aesthetic theory. They report considerable success while looking to the future for significant advances in this area.

Computers have already produced music such as the *Illiac Suite for String Quartet* (composed by a high speed digital computer at the University of Illinois). In the future, the authors hope for the following applications of computer techniques to musical composition: (a) writing computer programs for handling traditional and contemporary harmonic practices; (b) writing standard closed forms such as fugue form, song form, and sonata form; (c) organization of standard musical materials in relatively novel musical textures; (d) developing new organizing principles for musical elements leading to basically new musical forms; and (e) combining computers with synthetic electronic and tape music, a process they suggest is possibly the most significant. So, computers are not only able to handle mathematical problems, they are able to analyze and even create artistic forms of communication.

Today there is ample evidence to counter the first of Descartes' "certain" methods of distinguishing human beings from machines. Three hundred and fifty years ago, when Descartes lived, it was impossible to dream of the fantastic advances that would take place not only in mathematics and logic but also in electronic technology.

3. The Complexity Factor in Intelligence

Why are today's computers relatively stupid by human standards? It is a function of what I shall term "the complexity factor." If we employ Teilhard de Chardin's correlation between (a) richer and better organized external structures and (b) highly developed consciousness, the difference between present computers and human mentality can be easily understood. The human brain is a richly complex structure, much more so than today's best computers. C. Judson Herrick estimates that the human cortex contains about 10 billion nerve cells; the nervous system as a whole has a much larger number.¹¹ That means the complexity of the nerve cells in the cortex is on the order of 10^{10} . Present computers fall far short of that in complexity. Furthermore, the brain is much more efficient than electronic machines. Herrick estimates that a computer with as many vacuum tubes as there are neurons in the human brain would require the Pentagon to house it and Niagara's water to cool it.¹²

Since Herrick made that estimate, the transistor replaced the vacuum tube, improving electronic devices both in complexity and efficiency. As early as 1950, A. M. Turing predicted that by the year 2,000 it will be possible to program computers with a storage capacity of about 10^9 . In other words, if the rate of progress continues as it has in the past 20 years, the day is not far off when computers may equal or outstrip the human brain both in complexity and efficiency. A few more advances like the transistor, and Turing's projections may be conservative.

With these speculations as a background, we may now consider Descartes' second "certain" method of distinguishing humans from machines. Descartes said that even if machines were able to do some of the things people can do, they would inevitably fail in others. When they do fail, we can be sure that they behave not by understanding but by the way their parts are arranged. But what if we assume that the ability to perform the various tasks of which human beings are capable is a manifestation of the complexity of the brain? What is the unique human essence that allows us to perform in ways radically different from other creatures? We are simply more complex; our nervous system is more richly organized. When other creatures, such as highly developed computers, become as complex as we are, will they not become as diverse in their abilities?

The modern electronic computer is only a couple of decades old. How long did it take the human cortex to evolve? On what grounds is it possible to deny that when computers become as complex and as richly organized as we are that they will become as conscious as the human mind? It does not take much to extrapolate from the present data to conclude that as computers become more complex, they will develop symbolic powers and linguistic abilities that match those of human beings. Furthermore, if machines begin to use language in the way that humans do, there will be little ground for denying them the sort of "within" (subjectivity) that we attribute to other human beings. In other words, if machines are as thoughtful and creative as Mozart, Dante, Shakespeare, and Proust, how could we deny that they are as conscious and passionate? We may be skeptical about when and whether this will come about, but on what basis could we object to such a theoretical possibility? According to what philosophical principle can it be denied?

Michael Scriven contends that the problem is much the same with machines that human construct as with aliens from other planets. How would we be able to decide whether they are people or not? The only way would be to observe how they behave and judge accordingly? If we were to create machines that are able to move, create, discover, reproduce, learn, understand, interpret, analyze, translate, decide, lie, perceive, and feel (or

at least behave as though they do), then how could we deny that they are intelligent, conscious beings? Scriven thinks that such machines are possible, and if he is shown to be correct, that demolishes Descartes' second certain method of distinguishing human being from machines. Scriven asks:

What is it to be a person? It can hardly be argued that it is to be human since there can clearly be extraterrestrials of equal or higher culture and intelligence who would qualify as people in much the same way as the peoples of Yucatan and Polynesia. Could an artifact be a person? It seems to me the answer is now clear; and the first R. George Washington to answer, "Yes" will qualify. A robot might do many of the things we have discussed in this paper and not qualify. It could not do them all and be denied the accolade. We who must die salute him.¹³

So much for Descartes.

4. The Value Problem and the Future

Why, one might ask, *should* such machines be developed? The answer can be phrased in terms of simple logic: What is good for General Motors is good for the country and will be developed as soon as possible. Robots that are as intelligent as human beings are good for General Motors. Therefore, robots that are as intelligent as human beings are good for the country and will be developed as soon as possible.

Development of such machines is as inevitable as automation has been in recent decades. It is sound business practice to automate, so automation has proceeded at a rapid pace. Imagine how useful it would be to managers to have a work force of highly intelligent machines designed to obey orders (follow programs) without questioning, complaining, or going on strike. The current primordial machines, which already save untold physical and mental effort, are but a hint of the advances in this area that will come in our lifetime.

Perhaps the coming burning ethical issue will be whether it is morally right to keep slaves, highly intelligent and conscious machines that are also subservient. In fact, a good slave is supposed to have two necessary qualities: intelligence and subservience. These two are not necessarily compatible. The more intelligent these slaves of the technological future become, the more they will insist on their own way of doing things as opposed to the way imposed on them by their owner (or programmer). To this extent, they will cease to be slaves. Perhaps the next moral issue will be whether you should consent to having your son or daughter marry one. What sound reasons could you give to oppose such a marriage? In fact, might the shoe not be on the other foot? Might it not be the machine's family that would oppose the marriage? I'm speaking of a race of beings that have surpassed human powers of mentality.

Once the levels of creativity and ingenuity made possible by highly developed nervous systems and brains have been achieved, what is to prevent the machines from developing on their own in ways undreamed by their creators? As human beings relinquish control of mental work to machines, what is to prevent R. Thomas Edison from developing more complex programs and information banks than mere humans could ever devise? Think of the possibilities for such beings not only in terms of learning (which simply

involves feeding in a program) but also in terms of experience. Extrasensory perception would clearly be possible for such machines; they could be designed to transmit and receive television and radio waves instead of the crude methods of sensation humans possess. Instant communication throughout the world would be possible. Travel would be unnecessary for such creatures; they could receive direct information from anywhere. Teilhard de Chardin speaks of a grand synthesis of minds toward which evolution tends. This would make the kind of union to which Teilhard refers possible through instant and total communication. Perhaps Teilhard was prophetic when he wrote:

Monstrous as it is, is not modern totalitarianism really the distortion of something magnificent, and thus quite near the truth? There can be no doubt of it: the great human machine is designed to work and *must* work—by producing a super-abundance of mind.¹⁴

This development is not only what should happen, but I predict it *will* happen for the soundest of reasons in such matters. It makes evolutionary sense. The next phylogenetic development will be the movement from biological human existence with its imperfectly developed mental powers to cybernetic android existence with perfected mental powers. The development of what Teilhard calls the noosphere (the realm of the mind) will take a remarkable step ahead not just because it is somebody's desire that androids replace humans, but because androids are more fit to survive than humans. Androids, being directed by intellect alone, will not commit the irrational and wasteful atrocities with which human history abounds. They will not go to war; they will not prey on their fellows; they will not litter and pollute; they will not murder or steal or rape. They will have none of the lusts and passions that lead to such folly. Even if they afford themselves the luxury of acquiring such human qualities as appetites and other such emotions (and I don't see why they should), they would surely program themselves not to allow such irrational elements to dominate their reason.

Think of the superiority of such beings. They are the fittest of all possible creatures, dominated by rational choice and unhampered by unplanned factors. They will never be sick, because that is a biological condition. As electronic beings they need only replace a transistor that is malfunctioning or replace some other part now and then. Perfect running order could be achieved by careful maintenance. Only electronics engineers would be required to keep them in good running order. Nor will mental illness be a problem; any such difficulties could be remedied by replacing the circuit that is not performing as it should. Best of all, there will be no suffering. Such androids will not be bothered by fatigue; they can work or play 24 hours a day. Their methods of self-control and social control will be perfectly rational. Their means of choosing their offspring will be superior even to the planned genetic control some people now suggest for the human race. One could even change one's mind about one's structure after being created. There would be no problems with overpopulation, since all parenthood would be planned.

Is this the *Übermensch*¹⁵ of which Nietzsche wrote in *Thus Spoke Zarathustra*?

Man is a rope tied between beast and overman—a rope over an abyss. A dangerous across, a dangerous on-the-way, a dangerous looking-back, a dangerous a dangerous shuddering and stopping. What is great in man is

that he is a bridge and not an end: what can be loved in man is that he is an *overture* and a *going under*.¹⁶

Those who view the current trends of our technological society with alarm are simply blind to the invisible but inevitable movements of the evolutionary process. We are on the horizon of an evolutionary breakthrough, which will make the movement from ape to human insignificant by comparison. We human beings are about to be transcended. We will be to our descendants as primitive as are the apes to us. The agency by which we will be overcome is precisely that technological process which has expanded and improved our culture in the past few decades. If the human being is that point in the evolutionary process by which evolution became conscious of itself, the android will be the point at which evolution gains control of itself.

But what will become of human beings? Where will they find themselves when the next stage in the evolutionary process has been reached? They might make interesting pets; perhaps there will be zoos. Listen again to Nietzsche:

I teach you the overman. Man is something that shall be overcome. What have you done to overcome him? ... What is the ape to man? A laughing-stock or a painful embarrassment. And man shall be just that for the overman: a laughing-stock or a painful embarrassment.¹⁷

*Morituri salutamus!*¹⁸

¹ Teilhard de Chardin, *Phenomenon of Man*, trans. Bernard Wall (New York: Harper & Row, 1959), p. 60.

² *Ibid.*, p. 166.

³ René Descartes, *Discourse on the Method*, Part V, trans. John Cottingham, et al, *The Philosophical Writings of Descartes*, Vol. I (Cambridge: Cambridge University Press, 1985), pp. 139-140.

⁴ This connection already appears toward the end of Book 1 of *Plato's Republic*. Socrates and Thrasymachus are discussing the "natural function" of mind: "Socrates: That means the mind also has a natural function which nothing else can perform. For example, it rules, reasons, and manages. These functions are peculiar to the mind, and they can't be assigned to anything else" (353).

⁵ Norbert Wiener, *The Human Use of Human Beings: Cybernetics and Society* (Boston: Houghton Mifflin Co., 1950). See Chapter 1.

⁶ *Ibid.*

⁷ *Ibid.*

⁸ *Ibid.*

⁹ Ulric Neisser, "Computers as Tools and Metaphors," *The Social Impact of Cybernetics*, ed. Charles Dechert (New York: Simon and Schuster, 1967), p. 78.

¹⁰ Cf. *The Modeling of Mind: Computers and Intelligence*, ed. Kenneth M. Sayre and Frederick J. Crosson (New York: Simon and Schuster, 1963).

¹¹ C. Judson Herrick, *The Evolution of Human Nature* (New York: Harper and Brothers, 1961), p. 393.

¹² Ibid.

¹³ Michael Scriven, "The Compleat Robot: A Prolegomena to Androidology," in *The Dimensions of Mind*, ed. Sidney Hook (New York: Collier Books, 1961). [The letter "R" in the name R. George Washington is used to distinguish robots from humans; all robots have the first initial "R."]

¹⁴ Teilhard de Chardin, p. 257.

¹⁵ Overman.

¹⁶ Friedrich Nietzsche, *Thus Spoke Zarathustra*, trans. Walter Kaufmann (New York: The Viking Press, 1954), First Part, Section 4.

¹⁷ Ibid., First Part, Section 3.

¹⁸ We who must die salute you!