J. David Bolter

Artificial Intelligence

ARTIFICIAL INTELLIGENCE, writes Marvin Minsky, is "the science of making machines do things that would require intelligence if done by men."[1] Twenty years ago, Minsky and his colleagues were misfits in the world of computing. While others applied computing techniques to engineering and business, artificial intelligence specialists spoke of replacing the human mind, a "meat machine," with their more efficient electronic models, of creating nothing less than a new species for the planet. Today, work in artificial intelligence is more respectable, though still controversial. As our society has come to depend upon computers, it has grown less skeptical of the limitations of the machine. If twenty years ago there were only a few professors tinkering with programs to "rediscover" the Pythagorean theorem and to play poor chess, today they and their followers have established research centers in universities both here and abroad and are beginning to found companies.

That such work has indeed achieved legitimacy can be seen in the handsome support offered by the Department of Defense and the other customary sources of scientific funding. Even the pragmatic Japanese are turning to artificial intelligence as part of their assault on this country's hegemony in computer technology. All are lured by the promise that artificially intelligent computers will produce large profits and powerful new weapons. Robots will replace human workers at more complicated tasks than the assembly lines of today; smart bombs will find their targets with greater accuracy; programs will answer questions and obey commands given directly by corporate executives and military officers. In general, more of the operational planning so popular with business and the military will be performed entirely by computer.[2]

For years we have been reading in newspapers and popular magazines about the wonders of computers, and for decades, science fiction movies and books have featured all-knowing, though often malevolent, computers and robots. Many laymen believe that artificial intelligence is already with us, that some computers can understand the spoken word, read the newspaper, pilot a car using a television camera, or design their own even more talented offspring. In fact, no current program can do more than parody these achievements. Artificial intelligence is much harder to come by than the real thing: excellent programmers have been working for years on such problems and have made only a little progress. There is an enormous gap between what computers were built to do (mathematics and symbolic logic) and the wide range of skills that humans possess. Programmers must still work close to their machine's natural talents.

Computers are good at such well-defined and logical games as checkers, backgammon, and chess. They can play the first two about as well as any human, and can defeat any amateur at chess. For artificial intelligence programmers, however, playing chess and solving logical puzzles are still too mechanical and specialized. They want their machines to do something indisputably human, so they aim to endow their computers with the human facility for language; they write programs to read stories and news reports, "remember" the facts, and then answer questions about the reading. Or they tackle the problems of mechanical vision by devising a program to identify geometrical figures in a scene generated by a television camera. Automatic techniques for drawing inferences are used in their so-called expert systems, programs that depend upon hundreds or thousands of predefined rules-of-thumb: in a field of medicine, to diagnose patients; in geology, to suggest whether the particular rock formation is likely to contain oil; in chemistry, to identify organic compounds. Expert systems now operate in that socially uncomfortable limbo where the business world and the university meet. Meanwhile, business is making full use of robots, and the designers of commercial robots are drawing upon the research done by artificial intelligence specialists over the past twenty years. If anything, such robots have become too practical to merit the interest of many artificial intelligence specialists.

There are indeed practical applications for the techniques of artificial intelligence, but these applications are the less interesting half of the story. Despite its recent successes, artificial intelligence is still years, perhaps decades away from fulfilling its economic and military promise. Only scientific and industrial robots and highly specialized expert systems will have an important impact in the immediate future. For the rest, a breakthrough in computer chess will throw only a very small number of masters and grand masters into unemployment lines, and a coming glut of doctors in America may well save us from the need to install computers to diagnose our illnesses. Computers that can respond to simple English commands exist now and will certainly be improved. Looking beyond today's word processors, we shall soon regard computers as a full-fledged medium of verbal and visual communication. But it will be a long time before we can feed our home computers a copy of the *New York Times* and expect a reliable summary of the news.

Artificial intelligence is both compelling and controversial, not for its practical achievements, but rather for the metaphor that lies behind the programs: the idea that human beings should be seen as nature's digital computers. Ever since the prototypes were built in the late 1940s, computers have been invading domains that had previously belonged only to humans. They have solved differential equations for engineers and applied mathematicians. They have taken over such clerical activities as billing, inventory, and the printing of reports for the business world. They have taken control of machines in industrial processes. Before the advent of computers, these were tasks that required human intervention. Such mechanical devices as typewriters and adding machines could help, but men and women had to work in close association with these machines. The whole trend of the industrial revolution has been to make machines more self-sufficient, to move human controllers farther away from the process controlled. The trend began with the steam engine, if not centuries earlier with windmills, waterwheels, and the mechanical clock, and it continued with the power loom, the steamboat, the railroad, the dynamo, and the assembly line. The computer is the latest stage in the autonomy of the machine.[3] A programmer simply loads his code into the computer and sets it going. The machine may then act upon its instructions for minutes or hours, without further attention, and produce the most varied results in apparent independence of the programmer. One program may fly the space shuttle from orbit to landing with no help from the pilot; another may type out a letter in perfect paragraphs with underlining and boldface for emphasis. A program can even surprise its creator, often by committing some disastrous error. This ability to dispense with human control while performing complex tasks makes the computer metaphor almost irresistible to millions of professional and casual programmers. Is there not something human about this machine? Or is the reverse true? Is human nature itself mechanical, if capacities that seem characteristically human can be given to a machine?

This ambivalence has given rise to the project of artificial intelligence, whose goal is to achieve the complete assimilation of man and machine. If computers can already do mathematics, play games, and control other machines, artificial intelligence aims to endow them with every other function of mind and sense. Opponents of artificial intelligence, such as Herbert Dreyfus and Joseph Weizenbaum, find the computer metaphor absurd or dangerous, or both.[4] Proponents, including Nobel laureate Herbert Simon, enjoy the controversy, and have fanned the flames with extravagant claims of what their programs would be able to do in a few years. The decades have come and gone, and many of the claims have yet to be realized.

The most famous claim was made by Alan Turing.[5] A great logician, Turing was fascinated by the computer metaphor, and in 1950 predicted that by the end of the century, a computer would be able to deceive us with its electronic imitation of human faculties. He proposed the following test. Put someone at a computer terminal in one room and connect the terminal by wires to a second room. The person does not know who or what is in that second room: it may be another human seated at another terminal, or it may be a digital computer. To find out, he types questions into the terminal and receives answers from the second room. The interrogator is free to ask anything he likes. He can, for example, pose math problems, ask about English literature, or start an argument about politics. He has five minutes to decide whether there is a human at the other end of the wires or a machine. Turing predicted that by the year 2000, the computer would often succeed in fooling its interrogator. Now there is no computer or program at present that could participate meaningfully in the Turing test, and with only sixteen years left before the turn of the century, the prediction will probably fail. The terms of the test, however, are still quoted with approval by those working on artificial intelligence, and they remain faithful to their goal of proving that a computing machine can successfully imitate its creator.

In this respect, the artificial intelligence project is part of a long tradition.[6] For centuries, engineers and craftsmen have been making machines that imitate human beings, although their efforts have never been taken as seriously as the current project. The immediate forerunners of intelligent computer programs were electromechanical devices. In the fifties, for example, Grey Walter, a physiologist, designed an electromechanical "turtle" that could negotiate its way around obstacles on the floor. The turtle's technological ancestors include an electromechanical chessplayer built by the Spanish technologist Torres Quevedo. Before that, there were generations of fully mechanical automata. The most famous in the eighteenth century were those of Jacquet-Droz, including a boy who could write a message with pen and ink and a girl whose moving fingers could play a tune on a harpsichord, and a duck by Vaucanson that could quack and even eat and eliminate its food. The gardens of Europe in the seventeenth and eighteenth centuries were also adorned with figures powered by falling water: Neptune and Orpheus, and nymphs and shepherds from the pastoral tradition.[7] Moving figures, in fact, both human and animal, were represented on mechanical clocks as early as the thirteenth century, when these machines were invented in Europe.

Like the programs of artificial intelligence, all these hydraulic, mechanical, or electromechanical devices were attempts to imitate some aspect of human or animal nature. The artificial intelligence programmer would say that the earlier attempts were trivial because they were limited to external human features and rigid actions. That the writing boy of Jacquet-Droz looked human would be of no importance, since the boy could only repeat one or a few messages at a time and the messages had to be expressed as a complex interaction of gears. The effort was a tour de force, not a significant achievement of science or engineering. Gears were simply the wrong technology for expressing the fundamental qualities of human nature, such as problem-solving and the use of language. Even electromechanical devices could only point the way to the fully electronic technology that can express human nature.

There is some justice to this claim, because the computer *is* a more flexible machine than a clock or an electromechanical switchboard. Each technology, however, has had its own peculiar power as a metaphor. The idea of comparing the bodies of humans and animals to clockwork mechanisms was compelling to philosophers as important as Descartes and Leibniz. What both argued philosophically was illustrated mechanically by automata like those of Jacquet-Droz and Vaucanson. Artificial intelligence specialists play a similar role today. The toymakers were expressing the excitement of the precise mechanical technology of the day, an excitement that was finding a more practical expression in the machines of the industrial revolution. In the same way, artificial intelligence programs are illustrations of the possibilities of our new electronic technology.

ARTIFICIAL INTELLIGENCE AS A SCIENCE

Not surprisingly, artificial intelligence specialists do not see themselves as modern colleagues of the toymakers who amused the royalty of Europe in earlier centuries. Certainly, their work with computers requires a kind of training and intellectual rigor that toymaking did not. Our society is far more dependent upon science and technology today than at the dawn of the industrial revolution, and we accord technologists a higher status. Because their work has been questioned even by some other computer specialists, programmers in artificial intelligence are particularly concerned about status. They are toymakers who work with the expensive, powerful, and prestigious toys we call computers, and they want the appropriate recognition. In their writings, they try to present their field as a fledgling science: the science of cognition or the study of "thinking systems." They claim that only artificial intelligence can provide a precise and testable model for theories of memory, learning, language, and human inference. A good program, they say, is not merely a metaphor, an implied comparison between man and machine. Rather it is so much like a man or woman in important ways, that by studying its performance, we can learn about human thought or indeed about cognitive processes in general, of which computers and humans are both examples.

Load an artificial intelligence program into a computer and set it running. Of what is the computer now a model? The term "electronic brain" was once a common expression of the computer metaphor, but is the computer a model of the human brain? This was proposed in the early days of computing. Turing and another pioneer, John von Neumann, had thought of comparing the brain's "hardware" with the computer's. As early as the 1940s, Warren McCullough and Walter Pitts had described the mathematical properties of a net of neurons, drawing an analogy between electronic components and the human nerve cell. At the time, neurons still seemed to be fairly simple processors of electric information, and a digital computer might hope to reproduce the network of neurons in the brain. But the hope was never fulfilled. Although some elegant mathematics was discovered along the way, the elementary brain machines themselves were never more than toys. Scientists also began to realize that the brain, with its forests of dendritic connections and complex chemistry, was not a simple network of discrete logical components. In spite of the tantalizing fact that both computers and brains work their magic with electrical signals, the digital computer is not an adequate model of the human brain itself. Since the mid-1950s, most artificial intelligence specialists have shifted their efforts from physiology to psychology: their computer programs are now meant to be models of some facet of the human mind (memory, linguistic ability, inference) rather than the brain itself.

The artificial intelligence project has in fact borrowed the notion of a scientific model from other, more established sciences. However, computer models do not have the same status as those of physics and chemistry. The physicist's model is a series of mathematical equations and expressions, not tangible entities or metaphors. Often, scientists will devise metaphors to help them discuss the significance of the equations or to make their results more accessible to the layman. So, for example, electric current in a wire is compared to the flow of water through a pipe; the atom is said to be a miniature solar system, with electrons orbiting the nucleus like planets; Einstein's four-dimensional universe is sometimes compared to the surface of a three-dimensional globe. But for the physicist, the model is the mathematics and its interpretation; the comparison to some familiar physical object is of secondary importance. The wonder is that these highly abstract mathematical models should work in the world of experience, allowing scientists to predict and control events—that, as Galileo said, the book of nature should be written in the language of mathematics. The mystery of the mathematical prediction and control of nature remains a subject for philosophers of science. My point here is simply that the computer models of artificial intelligence are not like the mathematical models of physics. Computer models are not analytic in the same sense; they do not rely on deep mathematics, and they have no vitality and carry no intellectual conviction when separated from their machine.

Consider how the physicist and the artificial intelligence programmer each use the computer in their work. For the physicist, the equations and their interpretations form an abstract world that he can explore and modify with no more equipment than a pencil and paper. Often he may prefer to transform the equations into a computer program and see how the model performs with various assumptions and inputs. The artificial intelligence specialist may also begin his work in isolation from his machine. He may devise, with pencil and paper, an algorithm (a step-by-step procedure) to show how the mind stores, transforms, and retrieves units of information. He then translates this algorithm into a program, tests it on the computer, and adds refinements, always seeking to make the computer more faithful to his notion of how the mind works. But there is this crucial difference between the artificial intelligence specialist and the physicist: the former began his theorizing with the computer in mind. It may happen that a physicist cannot fit his theory into the computer if his equations are too complicated or of the wrong sort for direct computerization. This cannot happen to the artificial intelligence specialist, since, by definition, his model must be computable. In artificial intelligence nothing counts as an explanation of the human mind unless it is a step-by-step procedure operating on digitalized information. In other words, for the physicist, a computer may help to explore the model; for the artificial intelligence programmer, the computer (running the particular program) *is* the model. And nothing is duller than reading a prose description of a program for artificial intelligence. The program only comes to life in the machine, and then it is fascinating to watch—as fascinating for its mistakes, its confusions, and infelicities of language as for its successes.

The artificial intelligence project therefore depends upon a technological metaphor rather than a scientific model. It is possible to invoke the metaphor without a computer handy. In fact, psychologists today constantly speak metaphorically of visual "input" and linguistic "output," of the mind's strategy for "storing and retrieving" memories, of encoding and decoding messages. But behind these metaphors lies the machine itself. Much is made of our society's technological dependencies: on fertilizer, pesticides, and fossil fuels for agriculture, on railroads, automobiles, and airplanes for transportation, on titanium for national security. Here is a technology upon which a whole way of looking at the human mind now depends. Take away the computer, or (what is more likely) take away the excitement generated by this new technology, and the artificial intelligence project no longer convinces or threatens us.

SIMULATION AND REALITY

The artificially intelligent computer does not explain the human mind in the way that the physicist's equations explain the nature of atoms. Instead, it imitates the mind, and the point of imitation is to produce the same visible results without worrying about causes. Recall the definition of artificial intelligence as "the science of making machines do things that would require intelligence if done by men," a definition that stresses results over methods. The artificial intelligence programmer works by this pragmatic criterion: if a computer program acts like a human mind, then it deserves to be counted as one.

A philosopher with a concern for analyzing concepts and defining terms will not accept this pragmatism, and one such philosopher, J.R. Searle, has argued against the artificial intelligence project for just this reason.[8] Searle asks us to suppose that a programmer has in fact written a flawless program for understanding stories in Chinese. The program reads in the story in Chinese, accepts questions about its reading, and responds by printing sensible answers. All its input and output is in idiomatic Chinese. Like any program, this one works in a purely mechanical fashion, embodying the rules of Chinese chirography and grammar and applying those rules one at a time as it reads and responds. The program therefore reduces the ambiguity and complexity of the Chinese language to the purity of formal logic.

The artificial intelligence specialist would say that this program "understands" Chinese, for he accepts the definition that to know Chinese is to be able to transform input sentences into meaningful output. But Searle replies that the computer so programmed does not understand Chinese at all, and he proposes the following experiment by way of proof. Reverse the customary process of automation: replace the computer with a human being, one who speaks English and knows not a word of Chinese. Write out all the rules of the program in English, and let this human information processor apply these rules to sentences given him in Chinese. If necessary, have him memorize the whole procedure. Now this person can read stories and answer questions in Chinese, totally mechanically and by the way quite slowly, but he still cannot speak or write any Chinese on his own. He simply looks at patterns of lines on paper, consults his rules, and makes new patterns in reply. We would never say, Searle argues, that such an idiot savant understands Chinese, for such understanding is possessed by a human being as a part of his mental life, not as an externally imposed and memorized procedure.

Searle's human computer is a touching, rather Kafkaesque figure, who sits at his desk shuffling pieces of paper and recopying meaningless symbols according to methods he has learned by rote. Yet he has not convinced artificial intelligence programmers to abandon their work. They continue to believe that understanding Chinese can be explained by the formal logic of their machines. The dispute illustrates the utterly opposed points of view of the traditional philosopher and the programmer. For Searle is right by the standards of analytic philosophy. The program he imagines merely simulates a man answering questions in Chinese; it does not penetrate behind the visible effects to arrive at a theory of mental acts or intentionality. Searle skillfully reduces his argument to a point of common sense, when he writes with a note of exasperation:

> No one supposes that computer simulations of a five-alarm fire will burn the neighborhood down or that a computer simulation of a rainstorm will leave us all drenched. Why on earth would anyone suppose that a computer simulation of understanding actually understood anything? It is sometimes said that it would be frightfully hard to get computers to feel pain or fall in love, but love and pain are neither harder nor easier than cognition or anything else. For simulation, all you need is the right input and output and a program in the middle that transforms the former into the latter. That

is all the computer has for anything it does. To confuse simulation with duplication is the same mistake, whether it is pain, love, cognition, fires or rainstorms.[9]

Here, the artificial intelligence project has indeed abandoned common sense. In its excitement over the computer metaphor, it must abandon common sense, and indeed all computer programmers are tempted to do the same. Every one is inclined to confuse simulation with duplication because of the protean nature of the computer itself. This abstract machine can imitate the design, if not the physical effect, of any other machine and of many aspects of the natural world as well. The computer can simulate the operation of a jet engine, the traffic flow of a city, the reproductive rates of a colony of bacteria, or the invasion of Western Europe by Soviet tanks. Any phenomenon that can be divided into a series of discrete events can be simulated with some success by the computer. Every significant program is a simulation, an attempt to recreate within the computer some aspect of the world outside. Moreover, simulation encourages a particular view of the problem at hand, not the view of the mathematician who wants an exact and provable solution nor the view of the traditional philosopher. Instead, simulation is a matter of educated trial and error, of balancing one option against another, of minimizing evils and maximizing benefits. The programmer sets up his conditions and then tries out a series of possible futures: What happens to air quality in county X if we build a refinery in city Y? What happens to unemployment in city Y if we forgo the plant or if we build two plants? In reality, we cannot play with the future in this fashion, but in the world of the computer we can. We begin to regard the real world itself as simply one last run of the simulation.

The artificial intelligence project is an extension to the human mind of this intriguing aspect of the computer. On the one hand, the artificial intelligence programmer is quite practical: he simply wants a program that gives good Chinese answers to Chinese questions. On the other hand, he willfully ignores the difference between the "reality" of the world outside and the simulations inside his machine, between understanding Chinese and simulating that understanding. Computers are becoming better at manipulating elements within their tiny, simulated worlds, and more and more people are working with computers and coming to appreciate the style of work promoted by simulation, by trial and error, by thinking in terms of the process and its results rather than the deep causes. It may be that artificial intelligence will win this philosophical debate simply by sweeping away its opponents, refusing to engage them on their own ground. It will then offer its own answer to the great philosophical question: What is knowledge? The answer—that knowledge is the manipulation of symbols according to formal rules—will be wholly unsatisfactory to analytic philosophers and indeed to many philosophers of the older schools. The whole question will have been recast in operational terms provided by the computer. But of course the question has been recast many times in the past: by Socrates and Plato, by the Christian theologians, by Descartes and the mechanists. Now the technologists of the computer age, along with their allies in other disciplines, may have their turn.

PHILOSOPHICAL ENGINEERING

Artificial intelligence, then, is an exercise in symbolic logic, but it does not depend upon the relationship between mathematics and experiment that characterizes the physical sciences. Artificial intelligence has important links to modern philosophy and linguistics, but it is not philosophy in the sense of a careful analysis of the language we use to frame concepts. It is instead an intriguing combination of logic, technology, and philosophy—a combination that might be called "philosophical engineering." The oxymoron is intended. Engineers are eminently practical people, but the computer, the great achievement of electrical engineering, is also the most philosophical of machines. The computer turns engineers designing computing chips into philosophers of time, as they muse over the tiny fractions of seconds required for signals to thread their way along the chip. It turns programmers into logicians, as they try to decompose a mathematical or business problem into a step-by-step programmed solution. The ambiguity of the computer itself, a network of wires and transistors that somehow embodies theorems of logic, creates the ambiguous figure of the philosophical engineer. And workers in artificial intelligence are the most philosophical and ambiguous of all. Some may consider it a debasement to hand philosophy over to computer programmers; others seem to think the computer will revive philosophical debate. In either case, the development is significant.

The real homunculi created by the computer are not the chess programs and story understanders, but rather the computer architects, programmers, knowledge engineers, and artificial intelligence specialists themselves.

Man's view of his own nature has alway been conditioned in part by his contemporary technology. The best historical example is the effect of the clock and clockwork mechanisms on the mechanistic philosophies in the seventeenth and eighteenth centuries; but there are other instances from Greek times to the modern day in Europe, and no doubt in other cultures as well. In the past it was the philosophers who borrowed metaphors from technology: Plato and Aristotle from the Greek craftsman, or Descartes and Leibniz from the clockmakers. Today, thanks to their remarkable machines, the technologists have invaded the domain of the philosopher. The computer specialist dares to suggest that he can decide questions of epistemology, semantics, and psychology. The practice of "thinking through technology," as old as technology itself, has never yet become the dominant way of thinking in our culture. Yet this is precisely what the programmer in artificial intelligence is proposing: that we stop thinking of the computer as one modern machine among many and begin to see it as the true technological reflection of our own human nature.

MAPPING THE MIND

For my part, the most striking feature of the computerized view of man is this: that it reduces the mystery of human thought to a network for formal symbols. The computer reduces all problems to patterns of such symbols, often with great success. But the irony here is that formal logic, a product or at least a discovery of the human mind, is now being used by artificial intelligence specialists to explain the human mind itself.

In the computer metaphor, thinking is a process, the systematic progress of a computer racing through its program. But yesterday's thoughts, memories, and sensations must have some static representation in the machine, for they are data upon which our mental computer operates. Now, electronic data is composed of "binary units" or "bits" of information, which are merely strings of ones and zeroes. How can memories and thoughts be represented in such strings? The bits must be given a structure, a particular order and context that make them meaningful. Indeed, determining the appropriate data structure, deciding how the computer will interpret its ones and zeroes, is the programmer's main task. This is as true of the mathematician and the city planner as of the artificial intelligence specialist: they must all find suitable ways of arranging their data. However, the artificial intelligence programmer has the unusual task of finding data structures that will reflect the interplay of thoughts and sensations in the mind. He relies upon the computer's capacity to link together its discrete elements of data. In various ways, one string of bits in the machine can point to the location of another string, that string to other strings, and so on. The computer's memory can be organized like a road map, where the data elements are the towns and the pointers are the highways that lead from one town to the next. The trick for the artificial intelligence programmer seeking to imitate the mind is to choose the right combination of towns and highways, data elements and links among elements.

Here is an example. Suppose that our programmer wants to give his machine a knowledge of a common daily activity such as going shopping. He might start by sketching his structure on a sheet of paper. At the top of the sheet he draws a circle and writes "shopping" inside it. He then thinks of kinds of shopping: for food, for clothing, for an automobile, and the like. Each kind gets its own circle, and each circle is connected by a line back to the original circle. He proceeds to subdivide each kind. To shop for food, one may drive to the grocery store, get a basket, load it with vegetables, and so on. The circles begin to multiply, and the lines connecting them begin to crisscross. One may drive a car to the clothing store and to the drugstore as well as the supermarket, so that several lines converge here. The process of dividing actions into smaller subactions continues: to drive the car to the store, one starts the car, pulls out of the driveway, turns onto the main street, and so on. Suppose the programmer also wants to include the experiences that one particular person has had while shopping. These memories can be represented by more circles joined to appropriate places in the growing network. Now the programmer's paper has become a road map of the act of going shopping, in which the circles are activities or memories, and the intersecting lines are the associative or logical connections between these activities and memories. So far the structure exists only on paper. The next step, by no means trivial, is to replicate that structure inside the computer. Of course, the working programs of artificial intelligence are much more sophisticated about organizing their road maps than I have been. But they all rely on the principle of circles and lines, data elements and pointers, for there seems to be no other method by which a digital computer can process such information.

What is "inside" each circle? There is nothing more than a label, perhaps bits representing the word "shopping" or "auto" or "vegetable." Each circle is a formal symbol having no meaning in itself. In this computerized road map, meaning resides entirely in the interconnections between these symbols. And the meaning is realized only in operation, by programs that wander through the network of symbols—for example, a program that answers questions about what one may or may not buy in a grocery store. The thinking programs of artificial intelligence simply follow the links from one symbol to others, examining, modifying, and extending the network. I say "simply," but the programs are often ingenious, indeed triumphs of craftsmanship. They range from the coolly logical expert systems (the ones that diagnose patients) to programs that simulate the political views of a conservative voter. Artificial intelligence specialists may disagree violently over the appropriate structure, the choice of symbols, and the programming strategy. Some prefer a rigid hierarchy; others, a more diffuse pattern. Some prefer to crawl through networks of great complexity; others put more complexity in their programs and less in their data structures.

Nonetheless, in every program of artificial intelligence, the mind is a graph, and anything hidden or unverbalized simply slips through the interstices of that graph. Emotions and feelings can be part of the network as easily as logical decisions, but these, too, must have explicit links to other elements in the graph. Now, this is a revolutionary way of regarding the human mind. The computer and artificial intelligence are of course not solely responsible for the revolution: structuralism in linguistics and anthropology, behavioral psychology, and symbolic logic have played their part. The computer provides a technological focus for many lines of thought in the twentieth century, and the result is a psychology that cannot be dismissed lightly.

Consider for a moment the differences between the psychology of artificial intelligence and psychoanalysis, perhaps the most influential psychology of the last hundred years. Psychoanalysis is analysis in the old style: its seeks to probe beneath the surface, to find deep causes behind human actions. Its metaphor of the mind emphasizes the notion of depth, with the dark, instinctual portion of the mind, the id, buried below the more rational layers of ego and superego. The goal of psychoanalysis is to expose the repressed memories of childhood, and the psychoanalyst explains the human mind not as a processor of information but rather as a deep source of instinctual power and a shifting battlefield between passion and reason. There is a strong suggestion in Freud's writing that psychoanalysis is an unending process because the instincts, the id, can never be completely brought to light and rationalized.

There is an enormous contrast between this titanic, often pessimistic view of man and the view of the artificial intelligence specialist. A network of symbols in a computer program does not bring deep human motives to light, for there are no such depths. They disappear the moment we begin thinking of a human being as an information processor, a shuffler of symbols. The game of shuffling symbols may be tricky and indeed exasperating for the programmer, but the problems faced are wholly different from those of a psychiatrist, and not only because the psychiatrist is working with mental illness, with unusual or abnormal minds. In fact, artificial intelligence programmers sometimes try to simulate schizophrenia or paranoia, which they regard as special, perhaps pathological versions of information processing. Here, too, as in our shopping example, the flat, unambiguous network of symbols and the operational definition of success are their replacement for depth and causality in the representation of the human mind. Artificial intelligence specialists may even speak disparagingly of the idea of depth. Marvin Minsky has written: "To me 'intelligence' seems to denote little more than the complex of performances which we happen to respect, but do not understand. So it is, usually, with the question of 'depth' in mathematics. Once the proof of a theorem is really understood, its content seems to become trivial."[10] In fact, deep understanding in science or in psychology need not lessen the mystery at all. The more a physicist understands about subatomic particles, the stranger his world becomes, and the further he needs to look.

The more the Freudian psychologist probes the mind, the more he realizes the endless character of his task. But the specialist in artificial intelligence does have the experience Minsky describes: he does "reach bottom," almost immediately, as he maps complex human experiences into a data structure his programs can process.

* * *

The artificial intelligence specialists have, I think, gone too far. The computer is a mirror of human nature, just as any invention reflects to some extent the intellect and character of its inventor. But it is not a perfect mirror; it affects and perhaps distorts our gaze, magnifying certain human capacities (those most easily characterized as "information processing") and diminishing others. Nonetheless, the comparison of the human mind and the computer remains fascinating both for its distortions and its accuracies. We do not have to become religious converts to artificial intelligence in order to appreciate the computer metaphor. Nor do we have to join in the sterile debate over the question of whether computers can really think. Instead, we can ask in what ways the metaphor is apt and in what ways it may fail. Our view of the human mind changes from age to age, based upon social conventions, the work of poets and philosophers, and the current technical metaphors. Surely our contemporary task is to come to terms with the new electronic technology, a task that permits neither a complete rejection nor blind acceptance of the computer metaphor.

Predicting the future is so much a part of the project for artificial intelligence that I cannot close without making my own prediction. I think artificial intelligence will grow in importance as a way of looking at the human mind, regardless of the success of the programs themselves in imitating various aspects of human thought. It will color our view as long as computers themselves remain an important element in our technology. Eventually, however, the computer metaphor, like the computer itself, will simply be absorbed into our culture, and the artificial intelligence project will lose its messianic quality. We will have programs that perform various functions well and others less well—in natural language processing, expert systems, and in robotics. When, at some time in the future, the computer does lose its place as our leading technology, a new machine or technique will provide us with new metaphors, and the comparison of man and computer will become a topic in the history of science and philosophy. Essays on artificial intelligence (no doubt including this one) will then seem as quaint as *L'Homme-Machine* by the eighteenth-century philosophe La Mettrie, who created a scandal by arguing that human minds and bodies could be understood as clockwork mechanisms.

ENDNOTES

[1] *Semantic Information Processing,* edited by M. Minsky (Cambridge: MIT Press, 1968), p. v.

[2] A good example of the promises made by enthusiasts can be found in E.A. Feigenbaum and P. McCorduck, *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World* (Reading, Mass.: Addison-Wesley, 1983).

[3] The immediate forerunners of the computer were electromechanical devices, including the ingenious code-breaking machines designed by Alan Turing and others during the Second World War. See A. Hodges, *Alan Turing: The Enigma* (New York: Simon and Schuster, 1983).

[4] See H. Dreyfus, *What Computers Can't Do: A Critique of Artificial Reason* (New York: Harper and Row, 1972) and J. Weizenbaum, *Computer Power and Human Reason* (San Francisco: W.H. Freeman, 1976).

[5] In an article entitled "Computing Machinery and Intelligence," originally published in *Mind* and reprinted in *Computers and Thought,* edited by E.A. Feigenbaum and J. Feldman (New York: McGraw-Hill, 1973), pp. 11–35.

[6] Artificial intelligence enthusiasts are themselves aware of this tradition. For a summary that concentrates on the twentieth century, see P. McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (San Francisco: W.H. Freeman, 1979), pp. 3–90.

[7] For a description of some of these fascinating devices, see A. Chapuis and E. Droz, *Automata,* translated by Alec Reid (Neuchatel: Editions du Griffon, 1958).

[8] "Minds, Brains, and Programs," in *Mind Design: Philosophy, Psychology, Artificial Intelligence,* edited by J. Haugeland (Cambridge: MIT Press, 1981), pp. 282–306.

[9] Ibid., p. 302.

[10] "Steps toward Artificial Intelligence," *Proceedings of the IRE* 49 (1961): 27.