THINKING MACHINES

by Robert Wright

In July 1979, Italy's Luigi Villa, the world backgammon champion, took on a robot in a $5,000 winner-take-all match in Monte Carlo. The robot was linked by satellite to Pittsburgh's Carnegie-Mellon University, where a Digital Equipment Corporation PDP-10 computer, animated by a program called BKG 9.8, mulled things over. Villa was a 2 to 1 favorite; no machine had ever beaten a world champion in a board or card game.

But BKG 9.8 beat the odds. It won four of five games and, through judicious use of the doubling die, converted that advantage into a score of 7 to 1. "Only one thing marred the scene," recalled BKG 9.8's creator, Hans Berliner, writing in *Scientific American*. "Villa, who only a day earlier had reached the summit of his backgammon career in winning the world title, was disconsolate. I told him I was sorry it had happened and that we both knew he was really the better player."

Berliner's trade is that ambitious branch of computer science called artificial intelligence, or AI. Its goal, as defined by Berliner, is to make computers do things "that if a human being were to do them, he would be considered intelligent."

Defined this broadly, AI has room for two kinds of researchers. The field's "pragmatists" aim to replicate the results, but not necessarily the processes, of human cognition. They do not care if their machines *think* like humans, as long as they *act* like humans. Thus, the electronic chessboards that have brought automated defeat within reach of middle-income Americans do not win the way people win—by discerning and short-circuiting the opposition's strategies, or by forging boldly ahead with a master plan of their own, or by venting their aggression on a move-by-move basis. Rather, these machines rely on superhuman feats of calculation. At each juncture, they trace out thousands of possible sequences of moves and countermoves, noting the pieces won and lost, and then assign each possible action a number reflecting its likely long-term value. The rest even a human could do: make the move with the highest number.

The other kind of AI researchers are programmers who, like Berliner, see their mission partly as the duplication of the human thinking process. They write programs that work the way the mind works—or the way they suspect it works. To them, BKG 9.8

World War II gun directors, such as this one at a Newfoundland base in 1943, did more than help anti-aircraft weapons track enemy planes; they spurred early interest in the idea that machines could be imbued with intelligence.

represents a theory of how backgammon players think.

Whether or not programs such as BKG 9.8 can be said to show "intelligence," they have produced facsimiles reasonable enough to impress students of human behavior. AI has drawn the attention of cognitive psychologists in search of a fruitful metaphor for the mind, a fresh stock of terminology, or both. They have packed journals with "flow charts" of the human thinking process: Their models of the mind come complete with "preprocessing mechanisms" and "verbal protocols," and can "recover perceptual input"—even though they may labor under "incomplete feedback conditions."

As Princeton's George Miller has written, many psychologists have come to take for granted "that men and computers are merely two different species of a more abstract genus called 'information processing systems.'"

So have some journalists. The press regularly recounts the exploits of AI researchers whose progeny "think" like doctors and "understand" news articles. Alas, as computer scientists themselves concede, such accounts fall somewhere between oversimplification and distortion. *Newsweek,* reporting in 1980 that computers can "draw literary analogies" among Shakespearean plays, conjured up images of an IBM 4300 poring over *Macbeth* and then turning to a worn copy of *King Lear*. In fact, the computer scanned plot summaries that read more like the computer language FORTRAN than Elizabethan English: "Macbeth marry Lady-Macbeth. Lady-Macbeth is a woman—has property greedy ambitious. . . . Mac-duff is a noble—has property loyal angry. Weird-sisters is a hag group—has property old ugly weird—number 3."

Sticking to the Weather

Once the hyperbole is stripped away, computer scientists turn out to be only human—and to consider their machines only machines. AI's early optimism has been tempered. The difficulty of replicating even the more mundane cognitive functions has left some researchers saying what poets, mystics, and various other skeptics have said all along: The mind is not a computer. Putting it very bluntly, Marvin Minsky, former head of MIT's AI laboratory, says, "I'll bet the human brain is a kludge."

The field known today as artificial intelligence might well have been called "cybernetics," the rubric under which scientists first tried to simulate thinking electronically. Cybernetics began during the 1940s as the study of feedback systems. Its founder, the MIT mathematician Norbert Wiener, sought to make anti-aircraft guns self-aiming by giving them radar information about the speed and direction of targets. The parallels between this "feedback loop" and the human nervous system suggested that comparisons between mind and machine might be fruitful—an idea that fed on enthusiasm about new "electronic computing machines." Soon cyberneticists were building networks of elaborately interconnected switches, modeled after the brain's masses of neurons. But these "neural nets" displayed little intelligent behavior. By the late 1960s, this line of research had reached a dead end.

The term "artificial intelligence" was coined by Stanford's John McCarthy to describe a 1956 conference at Dartmouth, where he then taught mathematics. In a grant proposal submitted to the Rockefeller Foundation, McCarthy wrote that the meeting would address the "conjecture" that each aspect of intelligence can be "so precisely described that a machine can be made to simulate it."

The conference supported that conjecture. Allen Newell, J. C.

Robert Wright, 27, formerly an associate editor of The Wilson Quarterly, *is a senior editor of* The Sciences, *the magazine of the New York Academy of Sciences.*

Shaw, and Herbert Simon, three scientists connected with Carnegie Tech (now Carnegie-Mellon), together with the Rand Corporation introduced a computer program called LOGIC THEORIST. Confronted with 52 of the theorems proved by Alfred North Whitehead and Bertrand Russell in *Principia Mathematica* (1925), LOGIC THEORIST proved three-fourths of them—and one of its proofs was more "elegant" (i.e., straightforward) than the original.

Moreover, LOGIC THEORIST did not rely on brute force, trying every combination of logical rules until it found one that worked. Instead, it used "heuristics," rules of thumb that narrow one's focus in the face of numerous options that may lead nowhere. Newell, Shaw, and Simon, intent on modeling human thinking, made their program fallible.

Flushed with success, Simon ambitiously staked out AI's territory. There are now, he declared, "machines that think, that learn, and that create. Moreover, their ability to do these things is going to increase rapidly until the range of problems they can handle will be coextensive with the range to which the human mind has been applied."

Over the next few years, computer scientists produced one intriguing plaything after another. The Conversation Machine, built in 1959, could make passable small talk—so long as its partner communicated by typewriter keyboard and did not stray too far from the subject of the weather. In 1961, a program written by an MIT graduate student got an A on a calculus exam. By 1962, a string quartet had performed music composed by a computer that had used rules of counterpoint formulated by the 16th-century Italian Giovanni Palestrina.

The 'Common Sense' Problem

By the mid-1960s, though, the heady years were over. Impressive as AI's feats seemed, they still paled in comparison with the human mind's accomplishments. For example, General Problem Solver, a program unveiled by Newell, Shaw, and Simon in 1957, proved to be less capable than its name suggested. True, it was more of a Renaissance man than was LOGIC THEORIST: It could handle not only algebra problems but also logical puzzles, such as how to get three missionaries and three cannibals across a river alive using only a two-man boat. Still, these are not the kinds of skills most people associate with the word "generalist."

General Problem Solver's limitations suggested that intelligence cannot be boiled down to a few versatile techniques. It seemed, rather, that the human intellect depends on a large repertoire of tools, many of them useless without vast quantities of

PUTTING 'THE USELESS SCIENCE' TO WORK

ENIAC, the first fully electronic computer, blinked to life at the University of Pennsylvania in 1946. But the history of programmable machines goes back to Charles Babbage, the eccentric 19th-century English inventor of, among other things, the train cowcatcher. During the 1830s, he began work on his "analytical engine," which was to use steam power, punched cards, cogs, levers, and pulleys to solve mathematical and logical problems. Although the British government refused funds to build the contraption, its very concept raised the same machine-versus-man issues that the work of Artificial Intelligence (AI) advocates does today. Indeed, Babbage's collaborator, Lord Byron's science-minded daughter Ada, felt obliged to explain that, while the engine could do "whatever we know how to order it to" do, it had "no pretensions to *originate* anything."

The idea behind Babbage's machine (and ENIAC) originated a lot: an information industry whose worldwide revenues now total an estimated $175 billion and whose products are spreading to homes, offices, and factories everywhere in the industrial world. It has even spawned a genus of industrial robot that in 1982 numbered about 6,000 in America and 25,000 in Japan. Yet serious work on applications of AI, once called "the useless science," is fairly recent.

"Vision systems" are a high priority. Most factory robots must blindly follow their programmed directions; now ways are being developed for them to "see" and correct their errors as they go about cutting, welding, sorting, and assembling. Machine Intelligence Corporation of Sunnyvale, California, and Japan's Yaskawa together market a $105,000 "inspector" that compares parts on an assembly line with an image in its memory and removes parts that are bad.

Many firms are working on "expert" systems that can sift through a "data base" in a given field, answer questions, and offer advice. SRI International of Menlo Park, California, has stockpiled the expertise of geologists on natural resources in a program called Prospector. The program pinpointed a molybdenum deposit deep in Washington's Mount Tolman that had long eluded human prospectors.

Another AI goal has been to permit access to data-base information by way of plain English instead of requiring knowledge of some arcane computer language. Cognate Systems of New Haven, Connecticut, has designed a way of coupling a "natural language front end" with data on oil wells. To get, say, a map of all wells drilled by a cer-

specialized knowledge. Accordingly, during the late 1960s and early 1970s computer scientists turned their attention to "knowledge engineering," the transplanting of expertise from doctors, geologists, and mechanics to "expert systems." This research would eventually produce programs such as INTERNIST-I, an aid to medical diagnosticians: In a 1983 test tain firm in a certain area, an engineer need only ask for it. In another application, IBM is adapting an editing program called EPISTLE to summarize mail for busy executives.

To date, work on AI applications has been pursued mainly by small firms and academic researchers in the United States and Europe. This is changing.

In 1982, Japan, a laggard in the global computer sales competition, launched its first broad effort to develop "intelligent" products based on original, Japanese research. A joint venture of private firms and public laboratories, backed by a government commitment of $450 million over 10 years, it has been dubbed the Fifth-Generation Project, reflecting its focus on the new "massively parallel" computers intended to emulate human thought. (Computer generations are defined by their innards. Today's state-of-the-art machines—the fourth generation—are built around very large integrated circuits, called VLICs; the third generation used integrated circuits; the second used transistors; and the first, sired by ENIAC, had vacuum tubes.)

The Japanese, who describe their project as "the space shuttle of the knowledge world," aim to perfect a range of marketable devices, such as speech-activated typewriters, optical scanners that can read written language, and translating machines.

Britain and other European nations have launched major computer research programs. In the United States, still Number One in information technology, several computer firms have set up AI departments; 18 corporate giants, among them Control Data and Lockheed, have formed a research and development consortium, headquartered in Austin, Texas. But the big backer of advanced computer technology is the federal government, especially the Pentagon. In 1984, the Defense Department announced plans to spend $600 million over five years to develop new computer-based systems. While the focus is on military applications—such as a robot Army combat vehicle—the hope is to produce devices whose ability to see, speak, reason, and understand speech will have civilian uses as well.

U.S. spending by government and industry on advanced computer technology in 1984 alone may total $230 million. The stakes are high, too. Joseph P. Traub, head of computer-science studies at Columbia University, argues that progress in AI may determine which nation leads in computers during the 1990s—and, thereby, which "will be the dominant nation economically." Indeed, where might Britain be had it built Charles Babbage's analytical engine?

involving cases drawn from the *New England Journal of Medicine*, it proved nearly as accurate as the attending physicians.

But even with the mechanization of expertise, AI still faced the "common sense problem." Computers can play respectable chess and diagnose soybean-plant pathology with the assurance of a county agent, yet they cannot comprehend "The Farmer in the Dell."

"It's your home computer. It wants to know why you're not home." The rapid spread of low-cost "personal" computers, which first appeared in 1975, helped wire the notion of manlike machines into American popular culture.

In trying to imbue computers with common sense, researchers have had to grapple with questions of logic. How large a role does it really play in human thinking? How large a role should it play in machine thinking?

Marvin Minsky believes that the mind rarely functions with the rigor of logic: "I suspect we use it less for solving problems than we use it for explaining the solutions to other people and—much more important—to ourselves." Machines will not truly think, he suggests, until they can formulate vague definitions, harbor inconsistent ideas, and, on weighing evidence and finding it incomplete, jump to the nearest conclusion.

One of Minsky's favorite illustrations of logic's shortcomings is the "dead duck." Birds can fly, a duck is a bird, Joe is a duck. A computer with powers of deduction will conclude that Joe can fly. But what if Joe is dead? And what about Hubert the penguin, a bird who will never take wing? A child knows that neither can fly; a computer relying on deductive logic does not.

Exceptions can be programmed into a computer, but if there are too many it is not worth devising the rules in the first place. The real world, Minsky argues, is laced with both rules and exceptions, yet people cope anyway; deductive logic, therefore, must not be central to their thinking.

Researchers trying to teach machines to comprehend "natural language" (such as English) have confronted a second shortcoming of logic. Much of what humans absorb while reading does not follow logically from what is written. A newspaper reader does not have an airtight case in concluding that an assault victim who was "treated and released" was slightly injured. Still, such common sense reasoning is almost always on target.

Surviving Contradictions

Ambiguity further complicates matters. How is a computer to know that the meanings of *flies* and *like* change from one sentence (time flies like an arrow) to another (fruit flies like an apple)? Of course, context may clarify things. Is the computer at a college reunion or an exterminators' convention?

By giving computers such contextual information, Roger Schank, head of Yale's AI laboratory, has attacked several problems of language comprehension. Each of his "scripts" sets the context, providing generally safe assumptions about the way a given situation unfolds. Schank's "restaurant" script keeps the computer from even contemplating the possibility that "tip" refers to Gallant Prince in the seventh at Belmont, and also facilitates reading between the lines; when a customer leaves a big tip, the computer is told, it probably means that he liked the service.

Scripts are variations on "frames," a more general concept developed by Minsky. Both help computers cope with complexity by limiting the frame of reference to the situation at hand.

And, some researchers feel, both have limitations when taken as theories of human cognition. A single script or frame houses much information, but it would take a great many scripts to get a person through the day. Do humans really carry around thousands of separate frames and pop a new one into the mental projector every time they move from the food store to the street, or turn from the obituaries to the sports page? Is nature, with its preference for simplicity, really likely to build brains that have to perform such a complex juggling act? In their simplest form, theories based on frames suggest that this is indeed the case.

There are other theories of cognition that do not call for so much shuffling of information, but not all can be tested easily on conventional computers. They are more compatible with a coming generation of machines called "massively parallel," computers that some tout as the new wave in AI.

If machines are going to think like humans, Minsky says, they must quit defining words with mathematical precision and, instead, associate each word with a mélange of related words. They must be more like Euthyphro, the Greek sage who could name pious men but could not give Socrates a definition of piety.

"What if we built machines that weren't based on rigid definitions?" Minsky has written. "Wouldn't they just drown in paradox, equivocation, inconsistency? Relax! Most of what people 'know' already overflows with contradictions. We still survive." An "associationist" approach to defining words, he believes, will be easier with massively parallel computers.

Virtually all of today's computers are based on the "von Neumann architecture" developed by mathematician John von Neumann during the 1940s. A von Neumann machine is run by a central processing unit that retrieves information from the computer's memory, modifies it according to the program, and then either returns it to memory or prints it out and forgets it. Generally, such machines can do only one thing at a time.

In a machine with parallel architecture, though, different processors work on different aspects of a problem simultaneously. Though parallel computers have been around for some time, thus far none has been—well, massive. But Thinking Machines Corporation of Cambridge, Massachusetts, hopes to have a large prototype ready in 1985, and MIT is constructing a version of its own, the Connection Machine. Both will have some 250,000 processors, each powerful enough to be a capable computer in its own right; chips will be wired so that each one can communicate with any other. Even so, the machine will simulate only a thin slice of the mind, and MIT is already planning a larger version.

Majority Rule

In massively parallel computers, no one processor does anything very sophisticated, and none oversees the operation of the others. Intelligence is not imposed from the "top down"; it emerges from the "bottom up," much the way that collectively intelligent behavior arises in an ant colony despite its nonhierarchical structure and lack of individual genius.

Proponents of massive parallelism view the mind as a society. Jerome Feldman of the University of Rochester writes of "winner-take-all networks" in which "coalitions" of processors continually clash. In Feldman's model, concepts are represented not by strings of symbols, as in a von Neumann computer, but by patterns of interconnection among processors. This approach, he says, offers a way to address the issues of ambiguity and context more economically than do scripts and frames.

Take a sentence such as "John threw a ball for charity." In the machine envisioned by Feldman, the two senses of the verb *to throw*—to hurl, and to host—would live in separate processors, or "nodes." Upon encountering this sentence, both nodes would seek support for their interpretations; they would try to find other words in the sentence with which they have an affinity— with which they are connected.

Both would have immediate success. The *hurl* node is wired to the node housing the corresponding sense of *ball,* a spherical object. The second sense of *to throw,* to host, is linked with the second sense of *ball,* a dance. Once these two links are activated, they try to embrace one another.

Victory goes to the majority. When each pair tries to encompass the third key member of the sentence—the swing vote—only one succeeds. The *dance* node is connected to the *charity* node; charity balls are common enough to warrant that linkage. But the more conventional sense of *ball* searches in vain for a link with *charity.* The *host-dance* coalition now has control of the sentence and will electronically suppress any dissent.

In Feldman's model, as in models embodying scripts and

HAL grows up: In 2010, *the sequel to* 2001, *the ornery computer gives his "life" to save Capt. David Bowman (Keir Dullea) and spacecraft colleagues.*

frames, context helps. If "John threw a ball for charity" had come up at a social committee meeting, connections already activated would have headed off any grassroots drive for a *baseball* interpretation. Thus, Feldman says, the "connectionist paradigm" offers "dynamic" frames. They resolve ambiguity and take account of context, but do not come in bulky packages that must be juggled. Instead, a frame is defined by the prevailing pattern of interconnection among tiny packets of information, all of which stay put; dynamic frames can be modified subtly or dramatically without any reshuffling of information.

A Healthy Conflict

Ideas bearing some resemblance to Feldman's have been around for some time. In *Psychology* (1893), William James explored the "principles of connection" in accordance with which "points" of the brain are linked by "discharges" and thoughts "appear to sprout one out of the other." Later, came the cyberneticists' "neural nets," designed to learn by memorizing patterns of interconnection among nodes. Because neural nets did not live up to their billing, the von Neumann architecture was the only game in town by the 1960s, when psychologists turned for inspiration to computer science.

Almost every Psychology 101 student since then has encountered fruits of that search—textbook flow charts tracing the path of information through a mental processor and into long-term memory. Had massive parallelism been in vogue years ago, those charts might look different: Information might be dispersed through a huge honeycomb, and "bits" processed where they reside.

And the prospect of machines behaving intelligently might not seem so dehumanizing. No central processing unit will exert tyrannical rule over a massively parallel machine; the democratic behavior of the processors will be so unruly that not even a program's creator will always be able to predict results.

Would that uncertainty reflect a certain capriciousness on the part of the machine—even, perhaps, a trace of free will? Some computer scientists will go so far as to call such unpredictable behavior "nondeterministic"—which, in the language of philosophy, suggests freedom from mechanistic rules.

If massive parallelism lives up to the expectations of its strong advocates, this question may well be asked: Were the first 30 years of AI, with their emphasis on the "top down" approach to simulating intelligence, just a long detour for all the psychologists who were suckered onto the bandwagon?

Few in AI seem to think so. Whatever the value of massive parallelism as a metaphor for mind, no one contends that it can capture the entire thought process. Herbert Simon points out that, regardless of how information is processed at subconscious levels, it must pass through the "bottleneck" of conscious attention, which is clearly a "serial," not a parallel, processor; a person can entertain only one thought at a time.

Simon does not share Minsky and Feldman's high hopes for massive parallelism. He does agree that logic plays a limited role in thought—he won the 1978 Nobel Prize in economics for his theory of "bounded rationality," which stresses the arbitrary nature of much human decision-making. Still, he notes, conventional computers have shown an ability to simulate nonlogical processes, even if those simulations take longer than they would on parallel machines. Much enthusiasm about massive parallelism, he says, is "romanticism."

There is one point, though, on which massive parallelism's supporters and detractors agree: No matter which of AI's models of thought prevails, computer science will have made a lasting contribution to cognitive psychology. At the very least, computers force a theoretician to define his terms; it is hard to turn murky thinking into a successful computer program.

This benefit was foreseen nearly four decades ago by Harvard psychologist Edwin G. Boring. He had been challenged by Norbert Wiener to describe a capacity of the brain that no machine could ever duplicate. Just contemplating that challenge, Boring found, was enlightening. It forced him to refine his ideas about the nature of intelligence. Boring urged others to try this experiment in their heads—to pretend, in essence, that they were computer programmers trying to simulate human thought, and consider the issues that they would thereby confront.

In a 1946 edition of the *American Journal of Psychology*, he asked readers: "With what property must a robot be endowed by its maker in order that it may make discriminations, may react, may learn, may use symbolic processes, may have insight, may describe the nature of its own functions and processes?" Contemplating this question, he suggested, is "the way to go" at the question of how the mind works. "It is a procedure that keeps us clear."