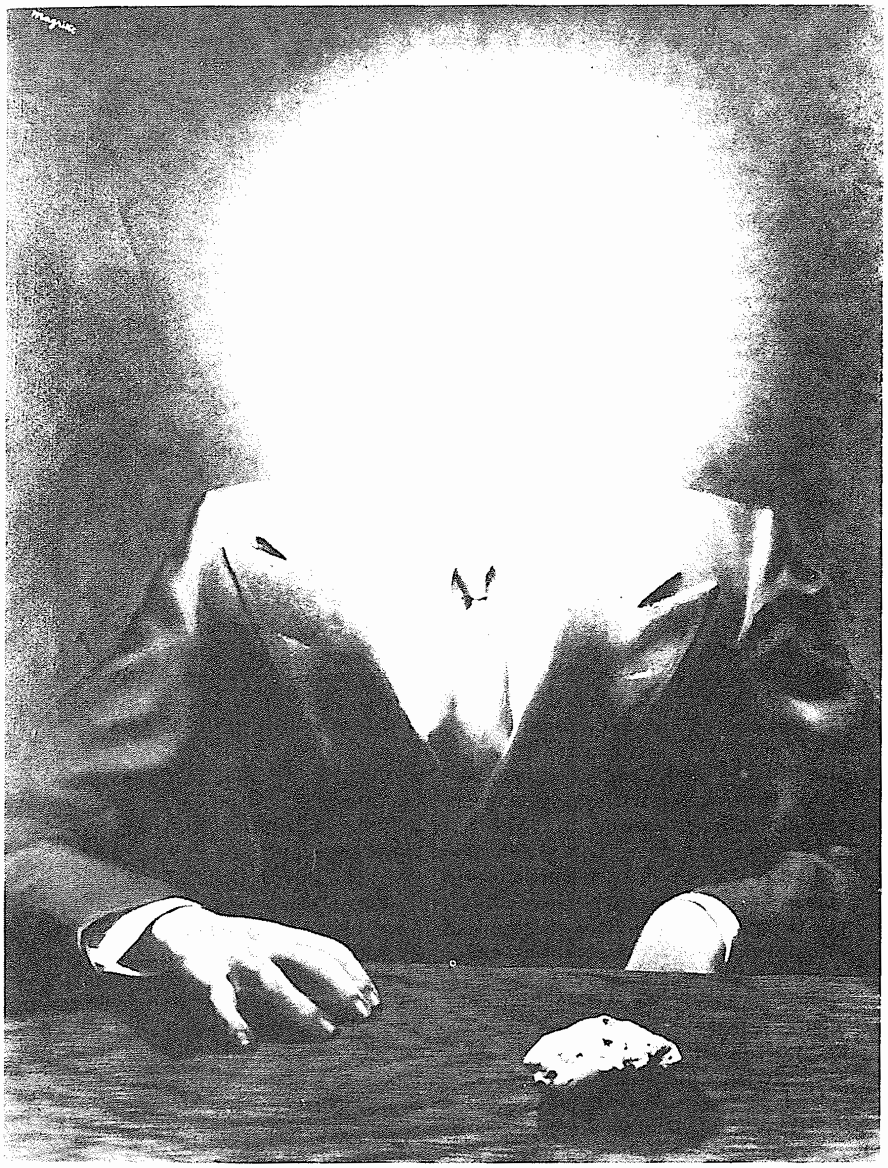


The Belgian painter René Magritte was a champion of surrealist art, which aimed at expressing the imagination as revealed in dreams, free of conscious control. Does a work such as his 1937 Le principe du plaisir *(The pleasure principle) suggest that the artist had a "mind"? If so, must such a "mind" be human? Could a computer have had Magritte's vision—and put it to canvas?*

The Mind

The first philosopher in the Western tradition was probably Thales of Miletus. Pondering the world around him in the 6th century B.C., he decided that everything is composed, in one way or another, of water. After Thales, things became more complex.

Later Greeks added other fundamental elements: air, earth, and fire. Then came Socrates' teacher, Anaxagoras. He argued that all the elements were directed by something even higher: *nous* (mind), the "purest of all things." Ever since, philosophers have puzzled over the mystery of the mind.

Does it exist? If so, how does it relate to the material world—including that mass of maybe a trillion nerve cells that is the human brain? Or is it a mirage, no more than an idea that *Homo sapiens* (Man the knower) developed in the prescientific era to explain the capacity for thought, feeling, and deliberate action that marks him off from the rest of nature?

Mirage or not, changing notions about the mind and the nature of reality have been important all through history. The foundations of religion, already weakened by the Reformation, were further shaken by Thomas Hobbes's argument that all is matter, the soul is a chimera, and free will an illusion ("nothing taketh a beginning from itself"). The growth of modern science both spurred and was spurred by the ideas of John Locke and other 17th-century Empiricists concerning the veracity of human perception. Metaphysical theories have shaped views about personal responsibility, and crime and punishment.

Even so, the mind faded as an object of academic curiosity in Europe and the United States some years ago. In the 1920s, psychology was claimed by the "behaviorists," who said that human actions can be studied solely in terms of stimulus and response, reflecting a Hobbesian view that people are basically automatons. Philosophy also left the mind off its agenda, especially after Gilbert Ryle blasted "the dogma of the Ghost in the Machine" in his 1949 polemic, *The Concept of Mind*.

Lately, the mind—or, rather, interest in the mysteries it represents—has made a comeback. On campus, philosophy courses are regaining popularity, and behaviorism is giving way to cognitive psychology, the study of the thinking processes. Meanwhile, new perspectives on the long closed world of mental operations are being provided both by advances in the neurosciences and the explosion in information technology. The question raised by HAL, the willful computer in the 1968 film *2001*, is getting serious attention: Can a machine have a mind?

Here, Richard M. Restak reviews the philosophers' struggle with the idea of the mind, Robert J. Sternberg explores the promise of cognitive psychology, and Robert Wright assesses the hopes that information technology may yield "artificial intelligence."

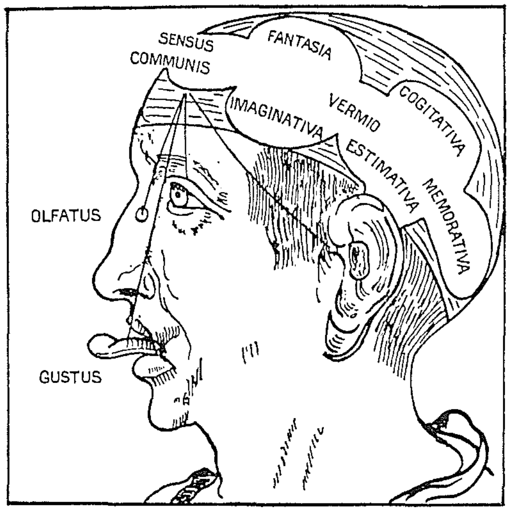
IS THIS CAT NECESSARY?

by Richard M. Restak

Someone once described a philosopher as a blind man in a dark room looking for a black cat that was not there. If the author was referring to a philosopher trying to define the human mind, he may have had a point.

At first glance, a definition of the mind seems obvious. After all, we speak of it daily. We talk of making up (or losing) one's mind, call some of our neighbors "mindless," and sometimes suggest that one of our nearest and dearest does not "know his own mind." But mostly, we use the word as shorthand for memory, feeling, intelligence, reason, perception, judgment, or something else. Do we add anything to our discussion when we speak of the mind instead of talking more specifically of, say, thinking or remembering? Or is the mind such a vague concept that, despite our best efforts, we are like the blind philosopher stumbling around a darkened room?

One reflection of our difficulty with the mind is the fact that there is no exact word for it in some languages—even German, the medium of many philosophers and of the founders of psychology. When Immanuel Kant was trying to create an anatomy of the mind for his *Critique of Pure Reason* (1781), he found that he could not even *invent* a precise term for the matrix within which, he claimed, are embedded sensibility, understanding, reason, and judgment. When they talk of the mind, Germans



The 16th-century Dominican friar Johannes Romberch saw the mind as a series of "faculties" that processed information received by the senses.

sometimes use the quaint term *Gemüt*, which refers to a person's nature. On other occasions they favor *Seele*, which corresponds to the Greek *psyche* and to soul in English. Then there is *Geist*, or spirit.

But none of these is quite right. Many who seek to understand the mind do not believe in a soul. And what exactly is spirit? Despairing of a satisfactory definition, some thinkers have decreed the black cat out of existence. In *Philosophy and the Mirror of Nature* (1979), the American philosopher Richard Rorty dismissed the mind as "just a blur—the sort of thing you get when you lay tracings of two delicate and complicated designs down on top of each other."

But if the mind does not exist, why was it necessary to invent it? And when did the invention take place?

No one will ever know if the idea first occurred to some cave man contemplating his image on the surface of a pond. But the earliest writing showing an awareness of something like what philosophers later called the mind is a series of "dream books" composed on clay tablets by the Assyrians in the fifth or sixth millenium B.C. These deal with dreams about death, the loss of teeth or hair, even the shame of finding oneself naked in public—all matters implying belief in a personal identity.

A society's view of dreams may be a measure of its sophistication about the mind: A belief in the reality of dream content implies a failure to distinguish fantasy from reality, without which a concept of mind is impossible. Primitive man, as philosopher Charles W. Morris noted in *Six Theories of Mind* (1932), "makes no sharp distinction between mind and nature, between a private and subjective life of consciousness and an outward world of corporeal events. There is no formulated problem as to how mind and nature can interact, or how mind can know a world that is not mind."

The ancient Egyptians' preoccupation with a god of dreams, Serapis, probably coincided with a concern about the relation of the body to the mind or spirit. Their observation that life depended on breath, and that death coincided with the cessation of respiration, provided the basis for a belief that the spirit dwells within the body but does not depend on it for existence. It was the spirit that required the food, jewels, games, and other items found in Egyptian tombs.

To the early Greek philosophers, the mind was pivotal in the maintenance of order and reason in the world. Anaxagoras declared it the ruler of "the whole revolving universe," which mind had put in motion "in the beginning." Plato went further, saying that the mind was not only immaterial but separate from the body, which it governed. But while the Greeks generally believed that the body was involved with the senses and conscious awareness, and that the mind's realm was knowledge, language, intelligence—and, most of all, reason—it was left to others to develop the problem that bedevils philosophy today.

The 'Third Eye'

That is the so-called mind-body problem. It dates from the 17th-century work of the first modern philosopher, the French scientist and mathematician René Descartes. Aiming to tear down the authoritarian ideas of medieval church philosophers and construct a basis for the advancement of science, Descartes argued in his *Meditations* (1641) that if all assumptions about reality are tested, there is only one assertion that is not open to doubt: that man thinks. The one proof of existence, he concluded, was consciousness: "I think, therefore I am."

Richard M. Restak, 42, is a neurologist and member of the faculties at Georgetown University Medical School and the Washington School of Psychiatry. Born in Wilmington, Delaware, he received an A.B. from Gettysburg College (1962) and an M.D. from Georgetown (1966). He is the author of Premeditated Man: Bioethics and Control of Future Human Life *(1975),* The Brain: The Last Frontier *(1979),* The Self-Seekers *(1982), and* The Brain *(1984).*

He said that the world was composed of both matter and immaterial spirit. Man's body hosts a "rational soul," or mind, which Descartes put in the pineal gland in the brain. Perception did not depend on the organ doing the perceiving. It was this "third eye" that evaluated events in and beyond the body.

In Descartes's day, when the brain's chemical and electrical workings were unknown, that was an arresting argument. He did not explain how matter and something immaterial could interact, in defiance of all known laws of nature. But his idea that somewhere humans have an inner observer who "experiences" and comments upon events in the surrounding world survived. This assertion of the existence of two interdependent but fundamentally incomparable aspects of reality, mind and matter, is known as dualism. Its legacy is a series of questions. Is the mind different from the brain? Can a computer have a mind? What is the role of consciousness in a theory of the mind?

The dualist view was taken to extremes by Bishop George Berkeley, the Irish-born, 18th-century Idealist philosopher. He maintained that the material world really does not exist: It simply consists of images in the mind, and God is the source of all perception. Said Berkeley: "To be is to be perceived."

Berkeley's Idealism was hard to come to terms with, as James Boswell's account of Samuel Johnson's riposte showed: "Striking his foot with mighty force against a large stone till he rebounded from it, 'I refute it *thus*!'" Of course, all that Johnson knew of the stone was his own experience of it—precisely Berkeley's point.

How, in fact, does one prove that a thing exists apart from somebody's perception of it? The British theologian and novelist Ronald Knox limned the problem in a pair of limericks:

There was a young man who said, "God
Must think it exceedingly odd
　　If he finds that this tree
　　Continues to be
When there's no one about in the Quad."

Dear Sir:
　　　　Your astonishment's odd.
I am always about in the Quad.
　　And that's why the tree
　　Will continue to be,
Since observed by
　　Yours faithfully,
　　God.

The Scotsman David Hume tried to resolve the mind-matter dilemma with an agnostic brand of Idealism called Skepticism. He argued that the mind was just "a bundle" of different perceptions, whose "ultimate cause is perfectly explicable by human reason." But in a moment of candor rare for a philosopher, Hume admitted continuing puzzlement: "I dine, I play backgammon, I converse and am merry with my friends; and when, after three or four hours' amusement, I return to these speculations, they appear so cold and strained and ridiculous that I cannot find in my heart to enter into them any further. . . ."

The Beetle in the Box

Well before Hume's time, the influence of religion was waning. Indeed, Berkeley's Idealism was meant to counter the efforts of various 17th- and 18th-century thinkers to expel the mind from the debate about reality. Spurred by such works as Thomas Hobbes's *Leviathan* (1651) and the ideas of the French philosopher Julien de La Mettrie, who viewed man as a soulless creature ruled by the laws of nature, materialism—the proposition that only matter and energy exist—took hold.

By the late 18th century, the German dramatist and scientist Johann Wolfgang von Goethe suggested an alternative to haggling over the mind's "nature." He argued that it is only truly revealed in action: "We exert ourselves in vain to describe the character of a human being; but assemble his actions, his deeds, and a picture of his character will confront us."

This emphasis on behavior has parallels in the Eastern philosophies, particularly Zen Buddhism, which takes the view that one is as one does. Goethe introduced in the West the idea that, as the late Walter Kaufmann of Princeton observed in *Discovering the Mind* (1980), "man is his deeds, that mind is what it does, and that the way to discover the mind is not through concept-mongering, but through experience."

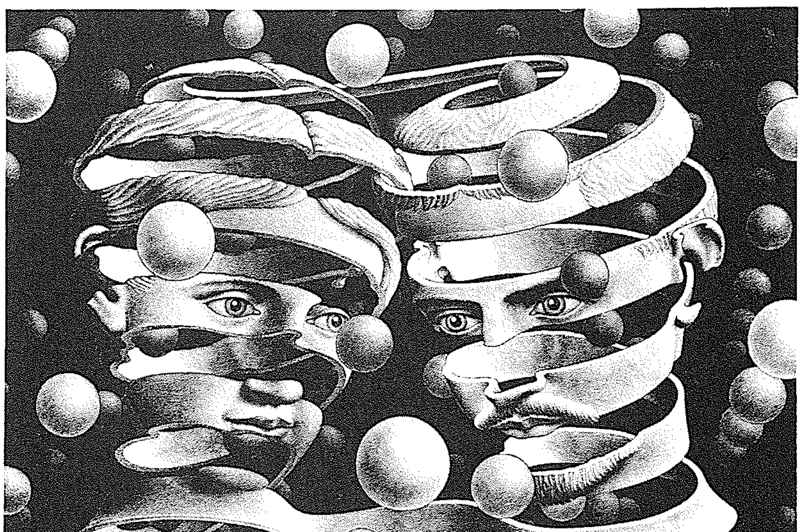
Experience implies development, a possibility ignored by earlier thinkers. In Kant's work, for instance, there is, as Kaufmann notes, "no inkling that the mind might change in the course of history, not to speak of biological evolution or the course of a person's life." The notion of a developmental mind would be central for later theorists, notably for the Germans who launched psychology in the 19th century and for Jean Piaget, the 20th-century Swiss student of child development.

If the mind develops over time, as such men believed, it is best understood as a process, rather than as an entity of some kind. The correct way to proceed, therefore, is to study feelings, emotions, desires, thoughts, and fantasies. The mind is all of these, yet it cannot be precisely defined by reference to any one of them alone.

To the psychiatrist, "mind" is synonymous with emotions; to the cognitive psychologist, it refers to the ways we perceive and process information; to the philosopher, it involves reason and logic. Each of these specialists focuses on a different aspect of man's mental life, and each one, like the proverbial blind men palpating the elephant, is convinced that his conclusions are the basis for broad generalizations.

Viewing the confusion raised by these approaches, some philosophers have contended that the existence of the mind is so *obvious* that no proof is required. Certainly it is difficult to disagree that there is something unique about what New York University philosopher Thomas Nagel calls the "subjective character of experience." If we wish to know the mind, we have merely to close our eyes and experience a state of utter subjectivity.

Yet when we try to describe what we experience by introspection, we run into trouble: Language cannot adequately express it. Moreover, what we learn cannot be verified in any way.



Philosophical issues fascinated the Dutch artist M. C. Escher. The figures in his 1956 lithograph Bond of Union *might be musing over a metaphysical question: Is the world—other people, even the universe—all in the mind?*

DEFENDING THE 'GHOST'

Just as most thinkers in the past saw the mind as a mysterious immaterial essence, most of those who ponder metaphysics today hew hard to some sort of materialism, a view that everything "mental" will eventually be explained in terms of physical laws. Two vigorous dissenters are Sir John Eccles, the Australian neurobiologist who won a Nobel Prize in 1963 for his work on showing how electrical impulses are transmitted in the brain, and Daniel N. Robinson, professor of psychology at Georgetown University.

In *The Wonder of Being Human* (1984), they argue that the most ardent materialists aim to "show that our ageless talk about the mind, feelings, and the like is but a vestige of religiomagical ignorances. It is finally 'ghost talk,' whose vocabulary will be properly translated by the findings of science, and thereupon eliminated from philosophically polite discourse." If the materialists have their way, they say, "all religions will finally be seen as the mythologies they are, and, apart from literary purposes, we will speak of the 'human condition' in the precise and morally neutral language of physiology."

What is new about this "nonsense," Eccles and Robinson maintain, is only "the willingness of otherwise sensible men and women to accept it." It was to underline his dismay about materialism in the 18th century, they say, that England's Bishop George Berkeley pressed his Idealism: the proposition that all one can really know of anything is what is in one's mind, such as a perception, an image, a thought, a memory—in short, an idea. Far from needing a material brain to have a mind, he argued, one first had to have a mind to know anything about matter. Berkeley's Idealism, say Eccles and Robinson, "was designed not to make us skeptical about the 'real world' but to show us how such a world is literally and factually unimaginable in the absence of mind."

Asked about Idealism today, modern materialists are apt first to dismiss it as "rubbish." But finally, the authors say, they are likely to make the same objection to it that modern dualists do, which is simply that it does not square with common sense: It "does not satisfy our deepest intuitive understandings of the relationship between ourselves and the world around us."

But materialism, the authors insist, fails the same test: Any view "that obliges us to deny the existence of thoughts, feelings, motives, will, memory, imagination, moral sensibility, and consciousness is false *because* it is incredible. For it to be incredible, there must be disbelief and, therefore, belief."

Thus, "arguments seeking to reduce mind to matter or to eliminate it altogether are self-defeating precisely because they are *arguments*." To argue is to believe. And to believe, Eccles and Robinson would have it, is to have a mind.

When someone looks into himself, the Austrian philosopher Ludwig Wittgenstein suggested, he encounters a "beetle in a box" that he alone can see. No one can enter into the consciousness of another and experience his "beetle." Introspection, as a proof of the mind's existence, is therefore a blind alley, bordered on one side by a vicious solipsism ("only *I* exist") and on the other by a hopeless agnosticism ("Who are you to tell me what I am feeling or thinking? Or I to tell you, for that matter?"). The existence or nonexistence of the mind thus becomes a puzzle worthy of Kafka.

Materialism and Machines

Attempts have been made to accept the situation, to say that the mind is simply an "immaterial substance." But does this make sense? Inflation, poverty, and health are not "substances" or "things" but, rather, terms to describe aspects of reality. In order to use the word *poverty*, most people do not mentally envision a ghostly substance that somehow exists outside of perceptual awareness. Yet when it comes to the mind, many thinkers persist in treating it as a specific entity, much as they might speak of a chair or a mountain range. What they are doing, however, is confusing a thing and a *process*.

This confusion permeates our language and culture. Colin Boakemore's *The Mechanics of the Mind* (1977) and Anthony Smith's *The Mind* (1984) are two popular books that are actually about brain research, neurophysiology. The titles imply that by studying the *brain* one can learn something about the *mind*—that it is something that can be taken apart, tinkered with, perhaps even tuned up.

The mind-as-mechanism idea persists even among scientists. In the 1950s, the British philosopher Charles Dunbar Broad observed that "if a man referred to his brother or to his cat as 'an ingenious mechanism,' we should know that he was either a fool or a physiologist. No one in practice treats himself or his fellow man or his pet animals as machines." Yet, he added, scientists "seem often to think it their duty to hold in theory what no one outside a lunatic asylum would accept in practice."

The mind-as-machine view has, nevertheless, received some support from studies of brain injuries. Damage to parts of the brain can produce mental impairment. Intelligence, memory, language, motivation, perception—all can be affected by an injury. Indeed, consciousness, basic to any theory of the mind, depends on a thumbnail-sized area in the brain stem; brain-surgery patients, who are usually kept awake during operations since no pain is felt when the brain is touched or cut, have lost consciousness when this area has been pierced by a probe.

It is thus tempting to declare that the mind is simply the brain and put an end to 2,000 years of speculative philosophy—despite the fact that the mind involves a series of activities that, now at least, cannot be explained in purely neurobiological terms. This declaration was made by the behaviorists, who decided that the mind was irrelevant and demanded that psychology restrict itself to what can be observed. While a generation of psychologists explored the workings of stimulus and response, "Skinner boxes," and maze-running rats, they ignored decision-making, changes of "heart," resolutions, religious conversions, and other mental operations that occur in the absence of any observable behavior.

By the 1960s, these and other weaknesses of behaviorism led materialist philosophers to new ways of explaining away the mind and mental phenomena. One argument was that, to understand the mind, one need only discuss matter and its "transformations" into sensations, images, memories, and other manifestations. Today, the most common expression of this position is the "central-state identity theory," which holds that the mind is nothing more than the state of the brain at any given instant. In this view, it is at least theoretically possible to infer thoughts from observations of changes within the brain, since the brain and the mind are identical.

This theory has several interesting implications. If the mind is indeed nothing more than a manifestation of one material substance (the brain), then there is nothing precluding the development of a mind within other material bodies (the silicon chips of a computer). To the committed materialist, when one achieves a sufficient degree of complexity in an "artificial-intelligence" machine—presto!—mind will emerge.

The 'Criteria Question'

The difficulty here stems from a persistent problem: the lack of firm criteria for defining what the mind is. If a machine achieves a mind-like status, how is it to be recognized?

John McCarthy of Stanford, a pioneer in the artificial-intelligence development effort, has this to say on the subject: "Machines as simple as home thermostats can be said to have beliefs, and having beliefs seems to be a characteristic of most machines capable of problem-solving performances." His point is that, since machines can be said to hold beliefs, a feature of the human mind, they too can be said to have minds.

René Descartes, here in a 17th-century engraving, believed that those who pursue philosophy should do nothing else: He sold the French estates he had inherited so as to be able to think "without cares or passions to trouble me."

Ponder the implications of this view. If thermostats possess beliefs, then the same could probably be said for automated doors, microwave ovens, surface-to-air missiles, burglar alarms, and other labor-saving mechanical devices—perhaps even toasters, which "decide" when the bread is done. The list of things with beliefs might finally include just about everything.

Which is no help at all. Instead of clarifying our ideas about mind, notes philosophy professor John R. Searle of the University of California, Berkeley, the thermostat-as-mind argument only confuses the issue. "For now the mind is everywhere. What we wanted to know is what *distinguishes* the mind from thermostats."

In recent years, several philosophers and artificial-intelligence specialists have tried to specify qualities that serve as criteria for defining the mind. Among the candidates, besides the holding of beliefs: purposeful behavior, "intentionality" (e.g., the ability to desire or choose), and a capacity for conscious, subjective experience. Some have suggested conscious awareness—though this falls afoul of the fact that when we are not conscious (as in dreamless sleep), we do not lose all of our beliefs and, say, cease to be Republicans or to desire world peace and universal brotherhood. Beliefs, or attributes such as jealousy, often operate outside of awareness, becoming conscious only at times.

Cognitive psychology and neurophysiology have only recently conferred scientific respectability on the insight that many important components of the mind remain permanently outside of awareness. In a word, we are denied *accessibility* to our own internal acts. Although it was Sigmund Freud who first popularized the notion, belief in *unconscious* actions is not at all limited to psychoanalysts. "A great deal of our thinking proceeds without conscious awareness," writes Oxford philosopher Stuart Hampshire. "In the exercise of the use of language itself and in many of our skills, we are thinking preconsciously, working things out without knowing how we worked them out, or by what steps we arrived at the conclusion."

The Great Mistake?

Consider an example cited by Bertrand Russell. "Suppose you are out walking on a wet day, and you see a puddle and avoid it. You are not likely to say to yourself: 'There is a puddle; it would be advisable not to step in it.' But if somebody said, 'Why did you suddenly step aside?' you would answer, 'Because I didn't wish to step into that puddle.' You know, retrospectively, that you had a visual perception, and you expressed this knowledge in words. But what would you have known, and in what sense, if your attention had not been called to the matter by the questioner?"

A proposed answer to the failings of consciousness and self-awareness as measures of the mind is to consider the mind as a form of information. In *God and the New Physics* (1983), physicist Paul Davies argues that "if the mind is basically 'organized information,' then the medium of expression of that information could be anything at all; it need not be a particular brain or indeed any brain. We are 'messages in circuitry,' and the message itself transcends the means of its expression."

Despite the appealing simplicity of this argument, it remains obvious that mental operations involve more than information. The essence of what we call the mind often consists of deciding what to do with information already at hand.

Machines, too, exhibit dualism: They are often engaged in activities that cannot be described in mechanical terms. "If a computer is carrying out mathematics," says Richard Gregory, professor of neuropsychology at the Brain and Perception Laboratory at England's University of Bristol, "you need mathematical concepts to describe what is going on; electronics is not adequate, mechanics is not adequate. So there's a dualism in a machine, but not a metaphysical dualism in the sense of a hovering mind affecting a machine. But rather that the procedures carried out by the machine are richer than and different from the mechanical and electronic processes described by the engineer, and I think this is also true of the brain."

This separation between physical structures and processes carried out by these structures applies to both machines and brains. The brain is a physical structure, yet it also carries out many processes—thinking, remembering, and the other activities that Descartes describes as *res cogitans*. What is gained, in the end, by bundling these processes and labeling the resulting mélange "the mind"?

The *invention* of the mind is not only unnecessary, but illogical as well. To introduce the mind when discussing neurophysiology or the mechanics of computers is to engage in what philosopher Gilbert Ryle, in his 1949 attack on dualism, *The Concept of Mind*, called a "category mistake": equating terms that actually are of different logical types. Contrasting mind and matter, Ryle suggested, is as illegitimate as would be the contrast of "'she came home in a flood of tears' and 'she came home in a sedan-chair.'"

To talk of the relationship of the mind to the brain, or the mind to a computer, is to make just such a category mistake. Thinking, remembering, reasoning, perceiving—such processes result from the activities of brains and, in recent times, machines as well. But these processes are not equivalent to a thing—the mind—any more than the flight of a swan can be considered a thing somehow separate from the swan itself.

Why postulate, as so many have, the existence of a "mind" interposed between a mysterious mental process and the brain, or computer, that makes such a process possible?

REINVENTING PSYCHOLOGY

by Robert J. Sternberg

Most of us spend much of each day thinking—about our work, the world we live in, and whatever comes to our attention. Cognitive psychologists are scientists who think about thinking itself. Can we identify the mental processes involved? If so, how do we use them? How might we improve them?

Serious people have been pondering the nature of thought for centuries. As a scientific pursuit, however, the study of thinking—cognition—is relatively new.

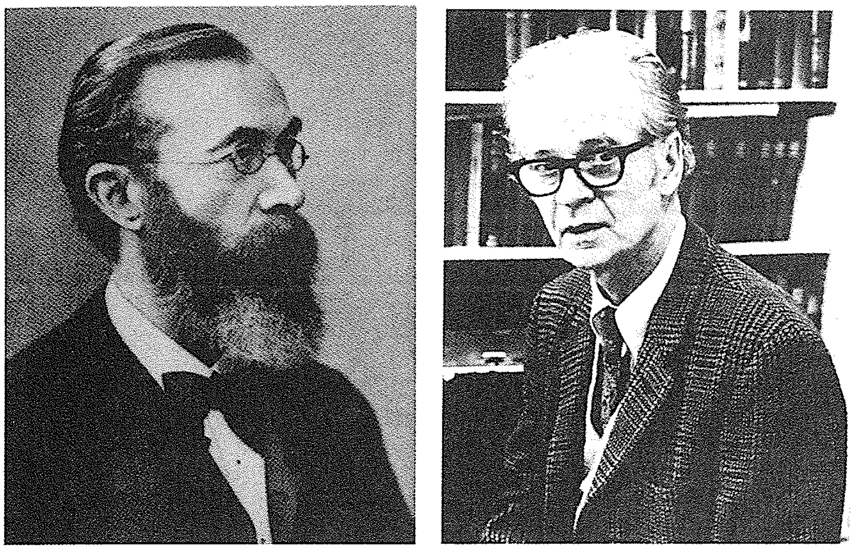
When Wilhelm Wundt, the German experimental psychologist, launched what he called the science of immediate experience during the late 19th century, scientific study of all kinds was flourishing. Physics, the most "objective" of all the disciplines, was probing the mysteries of matter and energy. But man, too, was a subject of wide interest, spurred in part by Charles Darwin's theory of evolution through natural selection. Physiologists approached the human body as an apparatus ruled by the laws of physics and chemistry. The psychologists, aiming to be as methodical as anyone else, also dealt with man as mechanism—*l'homme machine* as 18th-century philosopher Julien de La Mettrie put it. They focused on the observable and measurable. A distance was to be kept from what the founder of psychoanalysis, Sigmund Freud, would call "psychic reality."

But how to explore the mind's workings? In 1868, the Dutchman Frans Donders suggested that a start be made with a "subtraction method." For instance, he said, the time required to add two one-digit numbers could be found by subtracting the time it took to add four such numbers from the time needed to add five. Donders studied many mental operations in this way.

Modern cognitive psychology might have developed from that simple beginning, but it did not. Donders's work was attacked.

Critics argued that his method was scientifically invalid; there was a chance that subtraction itself altered the mental operation being studied. Enough psychologists shared this and other worries to abandon Donders's approach.

Psychology then took divergent paths. One was the stimulus-response approach championed by John B. Watson, in *Behaviorism* (1925), and further developed by Harvard's B. F. Skinner and others. To the behaviorists, mental processes were



The 19th-century German Wilhelm Wundt launched psychology as the study of mental processes; Harvard's B. F. Skinner and others shifted the focus to behavior, but now the science is again dealing with processes.

largely irrelevant. They aimed to explain behavior wholly in terms of punishment and reward, carrot and stick. This idea seemed to promise "results," and came to dominate psychology, especially in the United States and the Soviet Union.

The other path was the Gestalt, or "holistic," approach of the Germans Wolfgang Kohler and Max Wertheimer. They thought mental processes critical to organizing information provided by the senses, but hard to analyze. And even if one knew all the mind's processes, they said, one still would not understand mental performance well; the whole is greater than its parts.

The behaviorist and Gestalt camps debated for years over who best advanced the science. But by around 1960, the debate was stale. It was apparent that behaviorism just was not going to tell us much about how people handle complex functions, such as learning languages and solving problems, because it ignored the mental processes involved. Bothersome too was the behaviorists' denial of the existence of anything like free will—a fact that stirred novelist Arthur Koestler to blast behaviorism as "a monumental triviality that has sent psychology into a modern version of the Dark Ages." Holistic theory, for its part, seemed merely to redescribe mental phenomena, rather than explain them. Psychology was ready for a new approach.

As so often happens in science, a number of psychologists began to see things in a new way at the same time. By the late 1950s, Herbert Simon of Carnegie-Mellon University, George Miller at Harvard (now at Princeton), and others were urging renewed emphasis on cognition. The way to understand mental functioning, they said, was to understand mental processing.

By then, sophisticated research tools were available. The computer offered not only new ways of running tests, but also the possibility of simulating human cognition. Precision instruments improved experimentation. Saul Sternberg of Bell Laboratories was able to show how an alternative to Donders's century-old subtraction method could isolate mental processes without raising the doubts that Donders's tests had: In 1966, he measured operations taking as little as 40 milliseconds—the time required to compare a digit on paper with the mental representation of a digit in the head.

Thus equipped, psychologists were ready to probe cognition deeply. What kinds of things could they learn?

Consider this problem of analogy:

WASHINGTON is to ONE as LINCOLN is to
(a) FIVE (b) TEN (c) FIFTEEN (d) TWENTY

Analogies have long served as a basis for measuring intelligence. Early in this century, the study of mental ability was based chiefly on "factor analysis," a statistical method of relating intelligence-test scores to mental ability. If some students take tests in vocabulary, reading comprehension, figural analogies, and letter-series completions, those who test well (or poorly) in vocabulary might be expected to do the same in reading; ditto for figural analogies and letter series. This suggests that there are underlying verbal and reasoning "factors." Factor analysis has confirmed this.

But factor analysis could not take psychology very far. Tests could not prove one factoral theory of mind to be better than others, or even to be false; in science, a theory is not deemed worthy of attention unless it can be proven wrong if it is wrong. Factor analysis also just did not say much about the processes underlying intelligence: Merely to say that a good analogy-

Robert J. Sternberg, 34, is a professor of psychology at Yale. Born in Maplewood, N.J., he received a B.A. from Yale (1972) and a Ph.D. from Stanford (1975). He edited The Handbook of Human Intelligence *(1982) and is the author of* Beyond I.Q.: A Triarchic Theory of Human Intelligence *(1984).*

solver has strong reasoning ability means little.

Cognitive psychology deals with intelligence by separating reasoning into components. Faced with analogies such as the one above, people perform four main processes before they answer.

Encoding, which translates a stimulus into a mental representation. Here, one might encode the information that WASHINGTON was a president, is on a bank note, and was a war leader.

Inference, which finds a rule that relates the first term of an analogy to the second. The relations between WASHINGTON and ONE: He was the first president, is on a $1 bill, and was a leader in the first major American war.

Mapping, which finds a " higher order" rule relating the two halves of the analogy. Both WASHINGTON and LINCOLN are presidents, faces on bills, and war leaders.

Application, which generates a rule that forms a correct answer and rejects the alternatives. Here, the answer is (a) FIVE, reflecting LINCOLN's image on the $5 bill.

By analyzing processes in this manner, cognitive psychologists have addressed many questions. Some of the problems that have come under study:

▪ How long does thinking take? The time required by a mental process can be measured. For some verbal analogies, I have found that people tend to spend roughly 54 percent of their time encoding the terms, 12 percent inferring relations, 10 percent mapping higher order relations, seven percent applying relations, and 17 percent in giving the answer. If an analogy is solved in five seconds, 2.7 seconds would be spent in encoding.

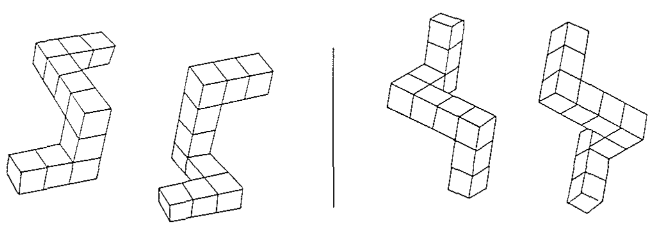
▪ What distinguishes "good" reasoners? In most inductive problems (those without a logically necessary solution), adept reasoners are usually faster than others at answering, but spend more time "up front" deciding what to do; in physics, for instance, experts tend to pause at the beginning of a problem to "represent" it with a diagram or a set of equations. Poorer reasoners are more likely to jump to a conclusion, then reach a dead end. They take a "local" approach, dealing more with the specifics of a problem than with its "global" aspects.

▪ What is "general" about general intelligence? People who are good at certain mental activities are good at others. Those who read with high comprehension tend to have big vocabularies, to be adept reasoners, to have large stores of general information, and to be articulate. To factor-analytic theorists, this suggested that a "general factor" of intelligence exists. But merely labeling "general" ability does not say what it is.

We have found that the mental processes cited above are used in almost all tests yielding a "general" intelligence factor. Research on analogies shows that while good reasoners tend to be faster than others in inference, mapping, application, and response, they are slower in the first step, encoding. They spend more time getting a clear sense of a problem and thus need less time to solve it. So general intelligence can be explained at least in part in terms of the processes used.

But even more important are the higher order processes that direct the operation. While inference and application, say, are important to general intelligence, even more basic is the "executive" process that decides to use them. Cognitive analysis has thus given us a solid basis for understanding general intelligence, which had been lacking before.

Psychologists are also studying the forms of mental representation on which mental processes act—the forms that information takes in the head. Much of this work grew out of a study published in 1971 by Roger Shepard and Jacqueline Metzler. They showed people pairs of perspective drawings like these:



The test participants had to judge as quickly as possible whether the figures differed only in rotation, or also in terms of a reflection. (In the pair at left, one figure is a mirror image of the other. The pair at right differ only in rotation.)

The key finding was that the time needed to recognize identical pairs depends on how far out of congruence they are. This suggests that people rotate images into congruence in their minds in the same way that objects can be manipulated in the real world. Other studies have shown that the speed of mental rotation—from 320 to 840 degrees per second—depends on the image. Letters and numbers can be rotated faster than figures. Robert V. Kail, Phillip Carter, and James Pellegrino reported in 1980, based on results of tests conducted at the University of Pittsburgh, that the speed with which mental rotation can be performed increases with age, at least from grade three to college. Stephen Kosslyn of Harvard has studied other aspects of imaging. When asked "Which is larger, an elephant or a fly?" people can answer rapidly. But if asked about a cow and a camel, most take longer; the smaller the disparity between the two objects, the slower the response. Other research has shown that in such tests people actually do seem to visualize the two objects and then compare their size.

Though not all psychologists believe that the case for such imaging is proven, most now seem persuaded that people *can* form mental images and move them around.

How Knowledge Is Gained

During the last decade or so, there has been much study of "domain-specific" skill and knowledge—that is, expertise. It is clear that we cannot fully understand excellent performance in any area unless we understand the role of experience.

The key study here was conducted by William Chase and Herbert Simon in 1973, with Master, Class A, and novice chess players. At the time, it was assumed that the experts had a "strategic" advantage: They could plan and "see" more moves ahead than others. But the Chase-Simon tests showed that, in fact, experts plan no further ahead than beginners (the intermediate players did the most forward planning). What marked the Masters was their experience: They could apply recollections of 10,000 or more board positions to their playing.

The findings about memory were also intriguing. The experts were better than others at recalling chess pieces in important board positions, but not at remembering other positions. They were adept at storing *crucial* facts.

From chess, research spread to other areas of expertise: reading, vocabulary, physics, medical diagnosis. Some of the most interesting work has been done on political problem solving by James Voss and his colleagues at the University of Pittsburgh.

For one study, they gathered four kinds of participants: political scientists specializing in the Soviet Union, political scientists with other specialties, political-science students, and chemists with no special knowledge of the Soviet Union. The participants were asked to imagine that they had been made head of the Soviet Ministry of Agriculture and now had to devise a plan to boost low crop production.

The more expertise the participants had, the more time they spent in setting up an initial representation of the problem. The chemists and students devoted the least time to this; the political scientists with non-Soviet expertise, more; and the Soviet specialists, the most of all. The results recall the tendency of

DID INTELLIGENCE END EVOLUTION?

Of all the things that differentiate man from other animals, none has been more important than intelligence, the ability to think and reason. Dolphins, elephants, and other species have developed larger brains. But only man has been able to use mental power to solve even elementary problems such as securing a steady food supply—a task that lesser creatures, with no agriculture, must face over and over again. How did man evolve, in only a few thousand years, from a simple hunter to a masterful being who can deal with complex matters just by, as Isaac Newton said of his approach to questions of physics, "constantly thinking unto them"?

The time and manner of the appearance of intelligence—now commonly viewed as a bundle of discrete mental abilities—is unknown, and may remain so. But in his book *Mind*, published in 1982 when he was at the University of Rochester, experimental psychologist David A. Taylor offers an intriguing hypothesis. He argues that man, in the course of evolution, developed the power to think in small steps that are analogous to those that children are known to take as they acquire the capacity to imagine, to communicate, and finally to reason analytically.

Like the fish, insects, and other lower order beings that first appeared 400 million years ago, a human infant does not think; it reacts instinctively to sensations, such as pain and hunger. Then, at about age two, it develops something known only to mammals and other higher order animals: the capacity to frame mental images, even without input from sight or the other senses. It can imagine things, drawing if necessary only on the information in its brain—a copious library that, by adulthood, can hold more than 500 times as much information as the *Encyclopaedia Britannica*.

This in itself was a useful adaptation: A person could imagine danger in places where he had encountered lions and avoid those places. But development went further: Man became able to imagine lions hunting, dozing, and doing other things. That is, he acquired the ability to produce mental images according to certain patterns. In much the same way that preschool children first begin to "think," humans learned to form sequences of images and to guide them according to learned rules. Research has shown that children do not truly begin to think abstractly until they master language, which trains them to order images in terms of rules. The same precept, Taylor argues, applies to the human race. Only after man acquired

good problem solvers to focus on global, up-front planning, but there is a difference: The chemists could be expected to be excellent problem solvers, but they did not do the global planning that the political scientists did. Yet research on chemists has shown that they do considerable global planning when faced language did he become capable of abstract thought.

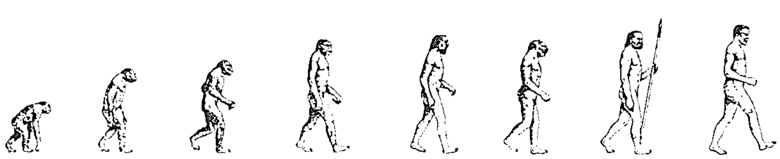
Like imaging, language was useful in itself: It "facilitated vital social practices such as communal living and organized hunting. The fact that it also made thinking possible may have been no more than a serendipity." But thought gave man his "tremendous" edge in the Darwinian fight for dominance in the world.

The final step in the evolution of intelligence, in Taylor's scenario, came when man started to think about thinking itself, including ways to improve it. This may have occurred as recently as the dawn of recorded history. Early writings show no sign of abstract thought or description of mental processes. Stories and myths displaying concern with the ideas and feelings of the characters involved came later—not too long before the Greeks began to study thinking and to lay the foundations of modern logic and philosophy. Thus, abstract thought is a fairly new phenomenon. Says Taylor: "The speed with which intelligence ultimately evolved once the conditions were right is testimony to the tremendous advantages it conveyed in the struggle for survival."

But in endowing man with intelligence, says Taylor, "evolution literally outsmarted itself." Through thought, man has reduced most of the usual threats to survival, such as hunger, disease, and climatic hazards. Result: "There is no longer any selection on the basis of genetic fitness; the weak as well as the strong survive to bear children, and there is no improvement from one generation to the next. In short, evolution appears to be over. . . ."

Biological evolution, that is. It has been replaced by cultural evolution, whose basic unit, "its equivalent of the gene, is the idea."

Among the products of this evolution has been the scientific method, which in a few hundred years has given man vast powers to shape his world. It may even let him "restart" biological evolution, through genetic engineering. In sum, Taylor argues, the advent of intelligence "brought several billion years of biological evolution to a halt," but man has "replaced it with a new form of evolution that is entirely under our control. We are, in effect, the inheritors of evolution. The future is ours to choose."



with problems in their own field.

The findings are convincing: Knowledge of a specialty plays a vital role in problem solving. An understanding of mental processes alone will not show how experts differ from others.

Yet all this has left some cognitive psychologists, including me, with a sense of discomfort. Certainly, experts know more than do novices, which is bound to lead to better performance: Nonexperts cannot spend much time in up-front planning if they have no knowledge of the field to apply. But one might wonder how the experts became experts and why others, with similar experience, did not. Not everyone who plays thousands of chess games will become a Master; not all who read a lot will become expert readers. To understand expertise, then, one must start not with knowledge, but with its acquisition.

Consider vocabulary. The view my collaborator, Janet Powell, and I have taken is that differences in vocabulary relate to differences in abilities to learn new words from their context. Try to define the two uncommon words in this passage:

> Two ill-dressed people—one a tired woman of middle years and the other a tense young man—sat around a fire where the common meal was almost ready. The mother, Tanith, peered at her son through the *oam* of the bubbling stew. It had been a long time since his last *ceilidh*, and Tobar had changed greatly; where once he had seemed all legs and clumsy joints, he now was well-formed. As they ate, Tobar told of his past year, re-creating for Tanith how he had wandered far in his quest to gain the skills he would need to be permitted to rejoin the company. Then, their brief *ceilidh* over, Tobar walked over to touch his mother's arm and left.

How do people figure out unknown words and thus build vocabulary? According to our theory, there are three important ingredients in the recipe for deriving word meanings: contextual clues, mediating variables, and cognitive processes.

Various contextual clues establish that *oam* means steam: We learn that the *oam* rises from a stew and that one can see through it. For *ceilidh* (reunion), we are given two temporal cues: that it had been a long time since Tobar's last *ceilidh* and that it is brief, suggesting that *ceilidhs* are rare and limited in duration.

Mediating variables affect our ability to use contextual cues. For instance, multiple appearances of a word (as with *ceilidh*) usually help us apply our cognitive processes to the cues. Three processes are critical here.

One is "selective encoding," by which one decides what information is relevant for finding a meaning. For *oam*, the cues are that it emanates from a stew, that one can peer through it, and that it is associated with fire. "Selective combination" enables one to assemble the cues. "Selective comparison" enables one to relate new knowledge to old knowledge. Here, one would consider things that relate to the clues—and come up with steam.

Selective comparison is especially critical in remembering new words. Often, one looks up a meaning in a dictionary, then soon forgets it. When one fails to relate a word to information one already has, it is difficult to retrieve later.

We have long known that vocabulary is the best single indicator of intelligence. But this did not make any particular sense in the absence of a theory of how some people acquire large vocabularies, while others do not. We now understand how differences in this, and in intelligence in general, can be traced in part to differences in ability to learn new words and concepts.

The processes of learning are not important only to vocabulary, of course. They also operate in what is known as insight. Consider some famous examples from science.

Alexander Fleming's 1928 discovery of penicillin was an insight of selective coding. In looking at a Petri dish containing a culture that had become moldy, Fleming noticed that bacteria near the mold had been destroyed, presumably by the mold. In essence, Fleming encoded this visual information in a selective way, zero-

Experiments with sound have raised hopes that IQ, now rated by written tests, might be measured by brain-wave activity. The waves triggered by aural stimuli have been found to be large and fast moving in bright people.

ing in on the part of what he saw that was relevant to the discovery of the antibiotic. He had no previously available cues for selective encoding to work on, but he focused on what to him was a new kind of cue—the destruction of the bacteria by the mold.

An example of an insight of selective comparison is Friedrich Kekule's 1865 discovery of the structure of the molecule of benzene fuel. After struggling with the matter to exhaustion, he slept and dreamed of a snake curling back on itself and biting its tail. When he woke up, he realized that the curled snake was a visual metaphor for the core of the molecule, which is a ring of carbon atoms.

Since we cannot probe insights of this caliber in experiments, my colleague Janet Davidson and I have studied more common ones—those needed to solve problems in such books as *Games for the Superintelligent*. Two examples:

▪ If you have black socks and brown socks in your drawer, mixed in a ratio of 4 to 5, how many socks will you have to take out to make sure of having a pair of the same color?

▪ Water lilies double in area every 24 hours. At summer's start there is one lily on a lake. It takes 60 days for the lake to be covered with lilies. On what day is it half covered?

Both problems require minor insights. People who fail the socks quiz tend to focus on the ratio of black to brown socks, and then to have trouble seeing how to use the information. But the ratio is irrelevant, as is seen by those who selectively encode that the only important facts are that there are two colors, and that a pair of the same color is needed. Even once this is encoded, one must selectively combine the information to realize that the answer is three socks; even if the first two one pulls out are brown and black, the third must make a pair.

The second problem also contains irrelevant information (that there is only one lily at first). It also requires selective combination to figure out that, with the daily doubling, the lake will be half covered on the 59th day—the day before it is fully covered.

Although people differ widely in their insight skills, research that Davidson and I have conducted shows that, to some degree, these skills can be acquired. After some weeks of drill in selective encoding, combination, and comparison, fourth, fifth, and sixth graders do better with simple insight problems.

Cognitive psychologists generally hope to use the knowledge they are gaining to improve people's thinking skills. Ultimately, many of us would like to see the day when what we are learning can be applied not only to everyday problem solving, but also to the thinking that policy-makers do when they make the judgments that affect us all.

Has cognitive psychology delivered on its promise? I believe that it has, in nearly all respects.

First, this science's aim was largely to find out what happens in one's head as one thinks. Though that goal got lost during the focus on behaviorism, research is now providing theories and methods needed for understanding mental processes.

Second, these theories and methods apply to interesting problems, such as the nature of imaging, insight, and vocabulary growth. Initially, with any new paradigm, there is a fear that it will answer only questions that no one cares about. Cognitive psychology has not had this problem.

Third, different aspects of what we are learning are coming together. In its early days, it seemed that the field might offer little more than detailed analyses of isolated mental operations without providing any understanding of how they relate to each other. This has not happened. For example, we have found that the insight processes are basically the same as those of vocabulary acquisition—though it is one thing to use selective encoding in divining the meaning of a new word, and another to apply it in finding that a mold (penicillin) is a potent antibiotic.

But no scientific approach is flawless. There is, in my opinion, one serious problem with cognitive psychology. It is *too* cognitive.

Thought is very much influenced by emotions, motivations, and desires. No matter how finely we analyze the thought processes and the mental representations on which they operate, we will not understand thought in its totality unless we understand how it is driven by, and drives, the noncognitive or "affective" side of human nature—love, pain, belief, will, and so on.

Cognitive scientists sometimes seem reluctant to acknowledge the need to combine their work with an understanding of these "softer" aspects of man's nature. Yet I suspect that we will never understand some of the most important decisions that people make, or the true reasons that they solve problems as they do, unless we probe the noncognitive as well as the cognitive side of the mind. This remains, as the philosopher and psychologist William James said of the nature of personality, "the most puzzling puzzle with which psychology has to deal."

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."

BACKGROUND BOOKS

THE MIND

In an ancient Indian legend illustrating the wisdom of the god Shiva, each of two men, a thinker and an athlete, has his head removed and grafted onto the other's body. The wife of each becomes confused as to which portion of her spouse she should stay with. Shiva, who sensed the importance of consciousness and knew where it lay, told them to go with the head.

Today, readers interested in the mind have a problem not unlike that of the wives. The literature is divided into two camps: There are writers who believe in some form of immaterial mind, and others who think that a material explanation of the brain will finally answer all questions about man's mental life.

Even good surveys of the field, such as **The Natural History of the Mind** (Dutton, 1979, cloth; Penguin, 1981, paper), inevitably take sides. Author Gordon Rattray Taylor leads his audience through the arcane mind-matter debate, with engaging side trips into anthropology and neurophysiology. But ultimately he concludes that "the great adventure of exploring the most complex system we know of in the universe" will justify faith in an immaterial mind.

Some who share this view are scientists. The most ardent dualist in print today may be the 81-year-old neurobiologist Sir John Eccles, whose many works include **The Self and Its Brain: An Argument for Interactionism** (Springer, 1977, cloth; Routledge & Kegan, 1984, paper), in collaboration with philosopher Sir Karl Popper, and **Mind and Brain** (International Cultural Foundation, 1982). Other dualist arguments are presented in the mathematician and philosopher Jacob Bronowski's **The Identity of Man** (Natural History, 1965, cloth; 1971, paper) and neurosurgeon Wilder Penfield's **The Mystery of the Mind** (Princeton, 1975, cloth & paper).

On the materialist side, the central testament remains philosopher Gilbert Ryle's **The Concept of Mind** (Barnes & Noble, 1949, cloth; Harper, 1983, paper). It was the first modern assault on dualism, whose tenets Ryle attacked with what he concedes is "deliberate abusiveness."

A rather more poetic early work of materialism is anthropologist Loren Eiseley's **The Mind As Nature** (Harper, 1962), which foreshadowed the "identity theory"—the idea that "mind" is simply the sum of what goes on in the central nervous system. More detailed treatments of the emergence of human awareness can be found in Gregory Bateson's **Mind and Nature: A Necessary Unity** (Dutton, 1979, cloth; Bantam, 1979, paper) and Julian Jaynes's **The Origin of Consciousness in the Breakdown of the Bicameral Mind** (Houghton, 1977, cloth; 1982, paper).

The demystification of the mind reached a peak in the branch of psychology that took its name from John B. Watson's **Behaviorism** (People's Institute, 1924, cloth; Norton, 1970, paper). Behavioral psychologists carried Watson's dictum that the study of human action "needs consciousness as little as do the sciences of chemistry and physics" as far as it would go. The idea that all behavior could be explained by responses to pleasure and pain was developed by Harvard's B. F. Skinner into an argument that personal liberty and free will (and thus good and evil) are just illusions. Skinner's **Beyond Freedom and Dignity** (Knopf, 1971, cloth; Bantam, 1972, paper) widened the gulf between the behaviorists and scholars with more "humanist" ideas.

Many of cognitive psychology's contributions to the study of mental operations are outlined in **The Mind's I: Fantasies and Reflections on the Self and Soul** (Basic, 1981, cloth; Bantam, 1982, paper) by Douglas Hofstadter and Daniel Dennet. Jerome Bruner's **In Search of Mind: Essays in Autobiography** (Harper, 1983), offers a broad view of what psychology has been able to determine about such processes as "knowing" and "learning," as well as about improving the intellect.

Herbert Simon lays out the hopes for *machine*-made intelligence in **Sciences of the Artificial** (MIT , 1969, cloth; 2nd ed., 1981, cloth & paper). Joseph Weizenbaum's **Computer Power and Human Reason: From Judgment to Calculation** (W. H. Freeman, 1976, cloth & paper) and Hubert L. Dreyfus's **What Computers Can't Do** (Harper, 1972, cloth; 1979, paper) suggests some of the limits to artificial intelligence (AI).

For those still uncomfortable with terms such as "parallel architecture," Pamela McCorduck's **Machines Who Think** (W. H. Freeman, 1979, cloth; 1981, paper) is a user-friendly history of AI. Those convinced enough by information technology to suspect that the mind indeed may be a mechanism—and a relatively poor one at that—may profit from Hofstadter's exuberant **Gödel, Escher, Bach: An Eternal Golden Braid** (Basic, 1979, cloth; Random, 1980, paper). His argument, spun out with engaging puzzles, riddles, and dialogues, is that the truly perfect thinking machine would, like man, be far from a creature of cold mathematical logic and precision.

The computer culture, rather than just the computer, is MIT sociologist Sherry Turkle's interest. In **The Second Self: Computers and the Human Spirit** (Simon & Schuster, 1984), she takes a careful look at the rising electronics in-group peopled by "hackers" and members of the AI priesthood. She concludes that the debate about "what computers can or cannot be made to do ignores what is most essential to AI as a culture: building not machines, but a new paradigm for thinking about people, thought, and reality."

The question of whether the mind is a machine may never be answered definitively, and maybe that is just as well. A Yes answer would devalue our sense of humanity; a No would deny our ability to understand ourselves scientifically.

Instead, we may be served best by a paradoxical conclusion: Yes, the mind is a machine; and No, it is not. The study of the mind can be approached profitably in both ways. Though this seems contradictory, there is an oft-quoted aphorism in physics: "The opposite of a shallow truth is a falsehood, but the opposite of a profound truth is often another profound truth."

—*Susan Baur*

EDITOR'S NOTE: *Susan Baur, 44, is a graduate student in psychology at Harvard. Readers may wish to consult books cited in the preceding essays, as well as in* WQ*'s previous articles on The Brain (Summer 1982) and Psychiatry in America (Autumn 1983).*