Artificial Minds

"...intelligence (whatever that may be) is more of a long term objective..."

Alex J. Champandard

The ultimate goal of artificial intelligence (AI) researchers is to create a computer program or machine that is capable of human intelligence—a difficult task, since no precise definition of intelligence exists. They do not neccessarily confine themselves to methods that are biologically observable, but their attempts at simulating or duplicating human thought have still provided invaluable insights for psychologists trying to understand the workings of the human mind.

Ideas about artificial intelligence (AI) have been around for thousands of years, ever since the Greek inventor and physicist Ctseibius of Alexandria invented a self-contained water clock—probably the first programmed machine—in about 250 B.C. It was a relatively simple device, but it inspired Greek philosophers to consider whether more complex machines could mimic humans. The oldest recorded examination of artificial life was a treatise by the Greek mathematician and inventor Hero (Heron), written in about A.D. 62. This work discussed automata, or robots, and was housed originally in the famous library at Alexandria, Egypt. It survives only in an Arabic translation, which is believed to have introduced changes to the original. All such ancient thoughts on artificial life and what constituted a

In the film AI, *intelligent androids perform many useful functions. When Haley Joel Osment is rejected by his surrogate family, he begins a quest to become a real boy, helped by "pleasure model" Jude Law.*

KEY TERMS

Cognitivism: The philosophical movement that led to cognitive psychology. It considers the mind to be an information processor with separate components, each of which can be studied in isolation. Models simplify complex phenomena by reducing them to smaller parts, so cognitivism is said to be reductionist.

Consciousness: A form of self-awareness defined as the rich, meaningful mental life characteristic of humanity. Some AI researchers see an evolutionary continuity to consciousness, meaning that it is present to a tiny degree in even the simplest of creatures. Others suggest that only humans have self-awareness.

Functionalism: The philosophical movement that considers the mind to be similar to a computer program. According to Turing's thesis (*see* box p. 147), functionalists would argue that a computer is capable of running a mind program, provided we can specify the program.

Infinite regression: The idea that consciousness exists only when internal meaning is perceived by something else internal—leading to an infinite number of theoretical entities that do this internal perceiving (*see* p. 153).

Intentionality: The capacity of a system to know the meaning of something outside itself: a key characteristic of human thought. For example, you can appreciate both the word “tree” and the concept it refers to in the world.

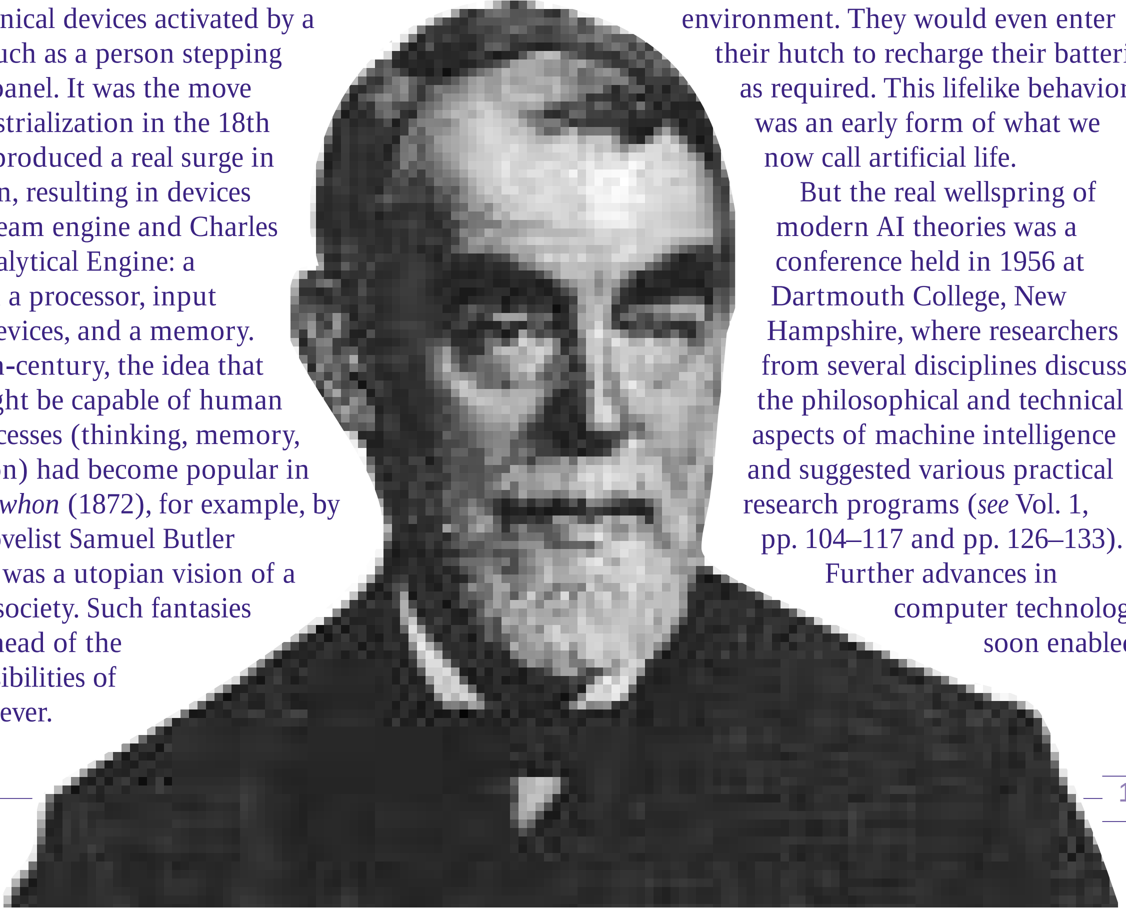
Intelligence: For AI researchers, intelligence is a system's capacity to show flexibility, understanding, and novel behaviors when faced with a problem. Most agree a fully human level of intelligence also requires consciousness.

Reductionism: A philosophical approach whereby a phenomenom can be explained from a viewpoint that is somehow simpler. For example, it explains complex mental phenomena in terms of simple properties that result from particular arrangements of nerve cells.

“mind” were purely theoretical (*see* box p. 143), but many of the concepts discussed are still applicable today.

By the 14th century simple automata had become relatively common. The gardening automata in Paris, France, for example, are reported to have aroused the interest of the philosopher René Descartes. These early machines had no suggestion of a mind, however; they were purely mechanical devices activated by a mechanism such as a person stepping on a hidden panel. It was the move towards industrialization in the 18th century that produced a real surge in mechanisation, resulting in devices such as the steam engine and Charles Babbage's Analytical Engine: a machine with a processor, input and output devices, and a memory.

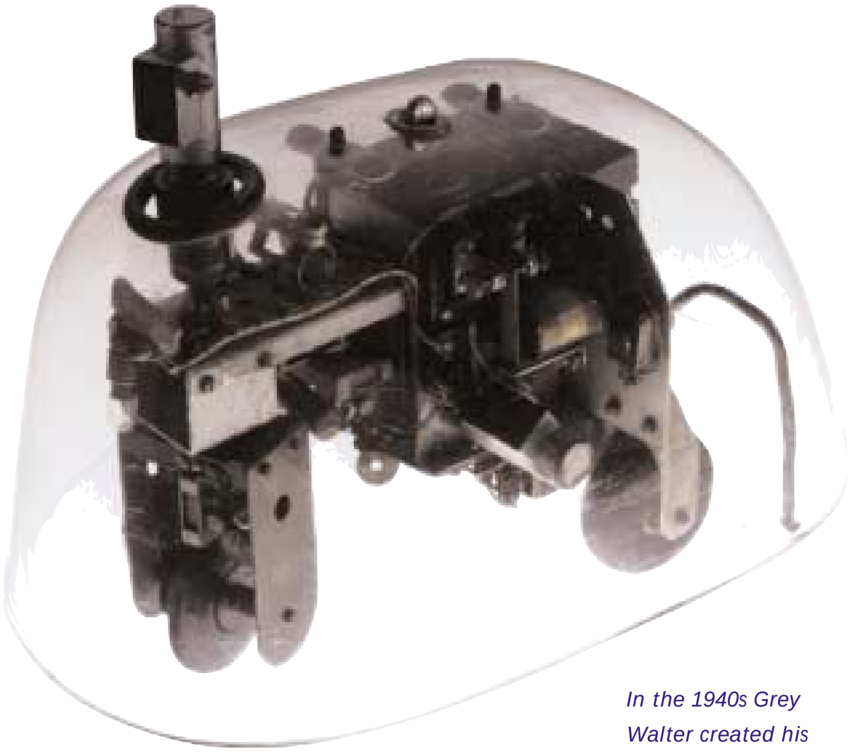
By the 19th-century, the idea that machines might be capable of human cognitive processes (thinking, memory, and perception) had become popular in literature. *Erewhon* (1872), for example, by the British novelist Samuel Butler (1835–1902), was a utopian vision of a machine-led society. Such fantasies still ran far ahead of the scientific possibilities of the time, however.



The novelist Samuel Butler wrote about the nature of human consciousness and of a society in which machines develop the ability to think: an idea popular at the end of the 19th century.

It was only in the 1940s that AI technology began to be developed, leading to early efforts at creating robots such as W. Grey Walter's turtles. These turtles were unique because, unlike earlier robotic creations, they didn't have a fixed behavior. Each turtle contained a touch sensor, light sensor, propulsion motor, steering motor, and a simple analog computer, enabling it to explore its environment. They would even enter their hutch to recharge their batteries as required. This lifelike behavior was an early form of what we now call artificial life.

But the real wellspring of modern AI theories was a conference held in 1956 at Dartmouth College, New Hampshire, where researchers from several disciplines discussed the philosophical and technical aspects of machine intelligence and suggested various practical research programs (*see* Vol. 1, pp. 104–117 and pp. 126–133). Further advances in computer technology soon enabled



In the 1940s Grey Walter created his Machina Speculatrix: an early form of artificial life. These mobile robot turtles were capable of complex behavior and were named for their speculative tendency to explore their environment.

researchers to simulate isolated functions of the brain, but the technology was still a long way from being able to create an artificial mind as capable, complex, and diverse as its human equivalent.

HOW DO WE MODEL A MIND?

In simple terms AI scientists attempt to create machines that can tackle some of the tasks accomplished routinely by the human mind, such as solving problems, playing games, recognizing speech, and seeing. But the mind is complicated and imperfectly understood and there is no universal definition of what constitutes intelligence. So how can researchers hope to copy what they cannot describe?

> ***"If we really understand a system we will be able to build it. Conversely, we can be sure that we do not fully understand the system until we have synthesized and demonstrated a working model."***
>
> ***— Carver Mead, 1989***

While it would be risky to assert that it will never be possible to replicate the human brain faithfully and accurately by artificial means, we should be aware of the complexity of the task. Inside your brain are about 100 billion neurons: nerve cells that carry information to and from the rest of the body and to each other. They can do so because they are connected in an overall pattern that determines intelligence, personality, memory, and so on. The adult brain has about 10 billion connections, but the number of possible arrangements has been said to be greater than the number of atoms in the universe. In other words, the number of possible brains is mind-bogglingly vast, and each one is highly complex. To build an artificial brain a

FOCUS ON

CHOMSKY'S CONTRIBUTION TO THE MIND DEBATE

In 1957 the linguist Noam Chomsky (*see* Vol. 1, pp. 118–125) published *Syntactic Structures.* Syntax is a term linguists use to describe the way that words are put together in sentences and phrases in both written and spoken language. He wrote the book in response to *Verbal Behavior* by B. F. Skinner (*see* Vol. 1, pp. 74–89), who believed that psychology was the science of behavior, not of mental events. Chomsky, however, asserted that mental events were vital to understanding how humans use language. He was also a rationalist because he believed that genes program certain mental representations: In other words he believed certain forms of human behavior and mental structure are innate.

Chomsky's book was highly influential in psychology and led to a movement away from behaviorism toward cognitivism. Cognitive psychologists see the mind as an information processor—and not just information in the usual sense of TV broadcasts or sport trivia, but all information, including sensory input. In other words they consider the human mind to be similar to a computer.

FOCUS ON

STUDIES OF THE MIND

Many different types of people study the mind, including philosophers, neuroscientists, AI researchers, robotics engineers, psychologists, and anthropologists.

Ancient Greek philosophers were among the first to record their thoughts about mental life (*see* Vol. 1, pp. 10–15). Plato (about 428–348 B.C.) believed that the most important knowledge was "already known" because people were born with it—a point of view known as rationalism. His student Aristotle (384–322 B.C.), on the other hand, argued that formal logic could be used to express knowledge but that logic itself was based on experience. For example, experience might yield the conclusion "all philosophers have beards." Aristotle's emphasis on direct observation of the natural world led to the development of empiricism.

Psychologists ask many of the same questions as philosophers but try to answer them with real-world tests rather than abstract theories—although not all of them use the same kinds of tests. Cognitive psychology is currently the dominant approach (*see* Vol. 1, pp. 104–117). It views the mind as similar to a computer and proposes that mental life consists of numerous, subtle mechanistic operations. Cognitive psychology is eclectic—it is rationalist insofar as it believes that people are born with mental systems for operating in the real world, and empiricist in that it believes that people are continuously collecting data for these mental systems.

Several ancient Greek philosophers recorded their thoughts on how the mind worked and how it processed information. This 17th-century portrait by Rembrandt shows Aristotle in Renaissance clothing, contemplating a bust of the epic Greek poet Homer.

scientist would have to reproduce at least some of this complexity and individuality.

The mind is also complex, but in a different way (*see* pp. 40–63). Some people argue that by understanding the brain you can understand the mind; others believe that the two must be dealt with separately because the brain is a physical entity while the mind is a philosophical concept, which some believe is equivalent to a soul. The brain is certainly capable of non-conscious tasks, such as controlling heart rate and breathing. Equally, the mind has qualities that are difficult to imagine happening physically, such as appreciating a work of art or picturing a far-away land.

Usually, those who study the mind see it as quite distinct from the brain, and tension between brain-level and mind-level explanations has characterized much of psychology in the 20th century. No one knows for sure which view is correct, but for pragmatic purposes the study of AI is based on the assumption that the mind is no more than a reflection of the physical structure of the brain.

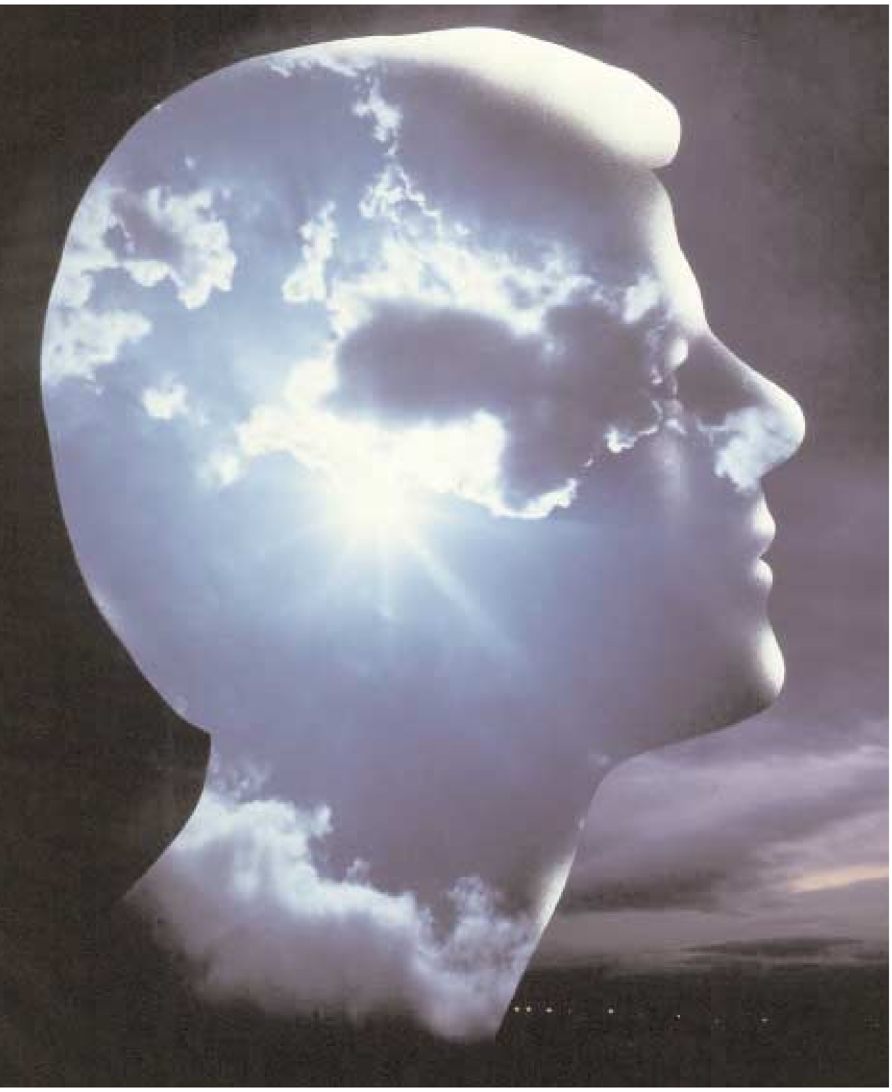
Practical approaches

So how do you begin to build an artificial mind? Science fiction may be full of them, but their creators do not have to roll-up their sleeves and actually make them (*see* box p. 146). Scientists who do can take one of two approaches. The first is to build a brain and hope that a mind emerges. The second is to analyze the ways in which a mind works, and then build a computer to simulate these workings.

Consider the first approach. Though it weighs an average of just 3lbs, the human brain is arguably the most complex object in the world (see pp. 20–39). It consists of multiple structures within structures, all requiring a blood supply, a body to hook up with, physical protection, an immune system, thousands of chemicals, and so on. To build one would require technologies far beyond our present reach.

The second approach tries to capture the essence of the mind, or at least those "parts" researchers believe are most important in producing human behavior. One advantage of this approach is that researchers do not need not to build the entire system. If they can discover certain parts and program them into a computer, they can see if their system works, then add more parts, and so on (a reductionist approach). They may even use methods that have not been observed in people or that involve more computing than the human brain can possibly undertake.

Many people believe that the mind is an abstract, philosophical concept that bears little relation to the physical structure of the human brain. This presents a real problem for AI researchers, for how can they hope to copy what they cannot define or describe?



Human limitations

Computers have plenty of speed and memory, but their abilities correspond only to the intellectual mechanisms that program designers understand well enough to put into them. Some abilities that children normally do not develop until they are teenagers may be put into programs, while some abilities possessed by two-year-olds may not. The matter is further complicated by the fact that the cognitive sciences have not succeeded in determining exactly what all our human abilities are. It is also likely that the intellectual mechanisms created in AI may be differ from those found in people. So whenever people perform better than computers on a task, or computers use a lot of computation to perform as well as people, this suggests that the program designers lack the necessary understanding of the intellectual mechanisms required to perform that task efficiently.

Despite these differences, AI has still provided many insights into the way that people think, shedding light on some of the strategies used by human minds and raising new questions about whether these strategies are indispensable or merely conventional. In other words, can thinking be done in only one way—the human way—or can the same conclusions be reached by different routes?

Modeling logic

Take logical argument. *The Oxford English Dictionary* defines logic as "a formal system using symbolic techniques and mathematical methods to establish truth-values in the physical sciences, in language, and in philosophical argument." In other words, it is an abstract way of representing relationships and inferring new ones.

The ancient Greeks were pioneers in this area, and one of the best-known and most widely used logical forms was the syllogism, two statements (premises) from which a conclusion could be reached by a process known as induction. For example: "All men are mortal; Aristotle is a man; therefore Aristotle is mortal." The final

FOCUS ON

CHESS COMPUTERS

One way of exploring the intelligence of a computer is by observing it in competition with a person—and for many years the focus of this area of research has been chess. Chess has specific properties that make it particularly well suited to the task: the game is usually played on a small area and has a set of rigid, clearly defined rules; at each stage of the game there are many possible future moves; and the winner must demonstrate a superior strategy to the loser. The last two properties are the most important because the vast number of moves available require the most successful programs to use strategies rather than simply computing every possible outcome of every possible move. It is this type of strategic problem solving that undoubtedly involves intelligence.

In the late 1960s M.I.T undergraduate Richard Greenblatt wrote a computer program called MacHack VI and entered it in a U.S. Chess Federation tournament, where it drew one game and lost four. He improved the program for the first American computer tournament of 1970 but it was beaten by Chess 3.0, a rival program from Northwestern University. Such early programs could "think" only two moves ahead. During the 1970s, however, computing power improved rapidly and by 1996 a dedicated computer-and-program combination named Deep Blue could think six moves ahead—whereas grand master Garry Kasparov of the former Soviet Union claimed he thought only three or four moves ahead.

If you are wondering why thinking ahead in chess is so difficult, consider the following. At the beginning of a game the white side can move 10 pieces. Eight of them are pawns, which can move either one or two squares forward. The remaining two pieces that can move are the knights, which each have two possible destinations, producing another four moves. So there are 20 possible moves that can be made. When it is the black side's turn, that player also has a choice of 20 moves. So the number of possible configurations after only one move in a game of chess is 20 x 20 = 400. For each new configuration there will be various possibilities for the next move—and if each move were the branch of a tree, each branch would grow its own tree with a similar or larger number of branches. As the pieces on the board become more isolated, there is greater freedom of movement, and the number of possible configurations becomes astronomical. In total, there are an estimated one million quintillion (10 followed by 23 zeros) possible chess games.

Many computer scientists have created chess-playing programs during their research into artificial minds. This is because the game requires the intelligence to think ahead and predict an opponent's next move.

A chess program can beat a person merely by number crunching, but it would find a grand master like Garry Kasparov much more difficult. That is because Kasparov plans his moves using mental short-cuts or strategies called *heuristics*. One such heuristic is "control the center of the board." By bearing the strategies for successful game play in mind, Kasparov can cut down the number of possible future configurations and make the task of predicting his opponent's moves much easier.

Intelligent use of specific rules such as these generally elevates humans above computers. In 1997, however, IBM presented Deep Blue, a chess-playing computer that could draw on a huge store of rules about chess-playing strategies. Kasparov, who was Chess World Champion at the time, played Deep Blue and lost.

statement goes beyond the information contained in the two premises but will be true as long as the two premises are true.

The syllogism is a powerful form of argument, and many people use it much of the time. But is it necessary—is it a crucial part of human reasoning? The evidence is inconclusive, but many people find it hard to solve certain problems that should be easy with logical analysis (*see* Vol. 1, pp. 134–143). Thus it seems that logic is not entirely natural—it is learned,

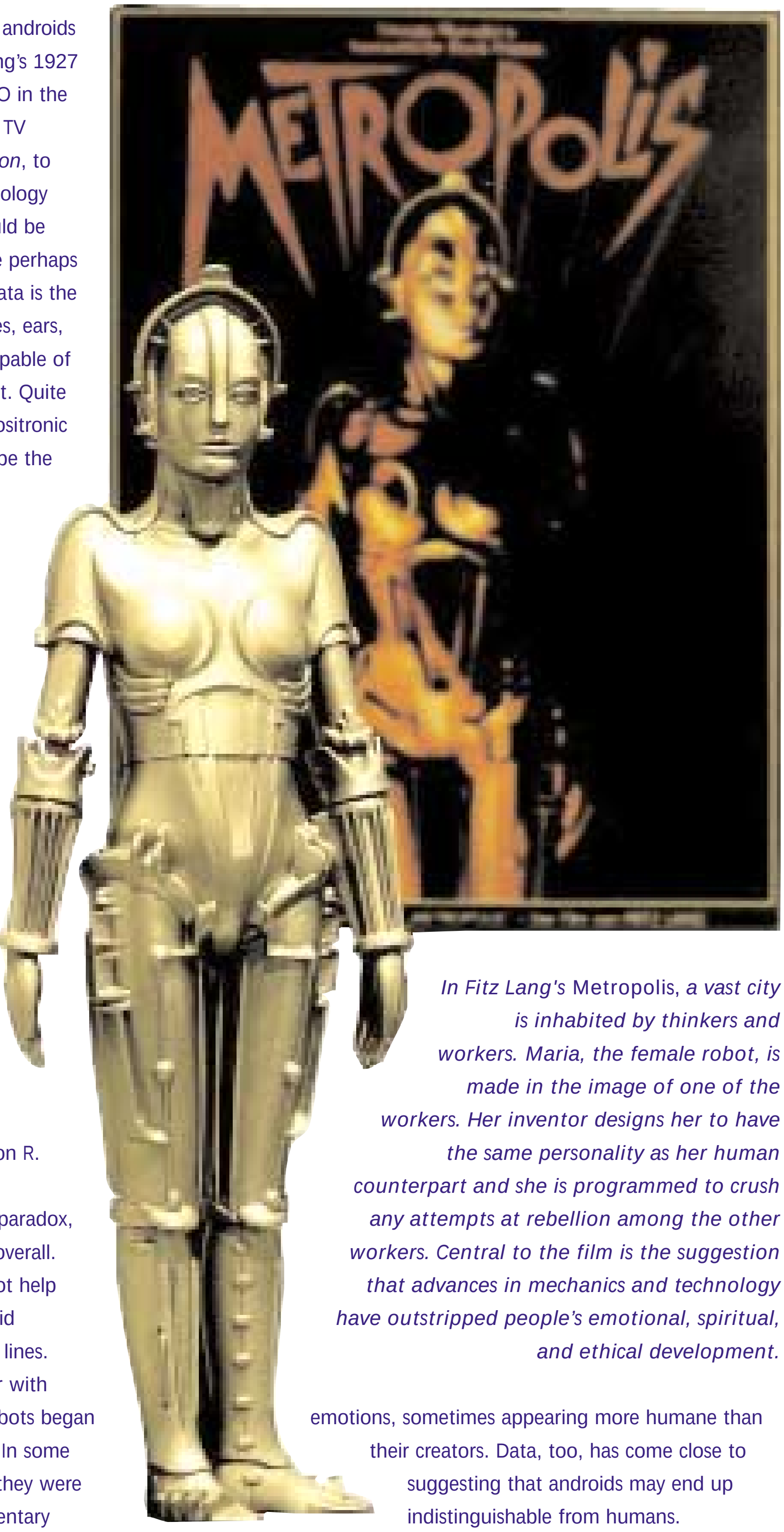
FOCUS ON

FICTIONAL ROBOTS OF THE MODERN AGE

Science fiction is full of robots and androids—the female archetype in Fritz Lang's 1927 movie *Metropolis*, R2-D2 and C3PO in the *Star Wars* movies, and Data in the TV series *Star Trek: The Next Generation*, to name but a few. If built, the technology involved in their development would be immense. The *Star Wars* droids are perhaps the most physically crude, while Data is the most complicated. He has hair, eyes, ears, and human-looking skin, and is capable of all the human ranges of movement. Quite apart from his Asimov-inspired "positronic brain," if Data was real he would be the eighth wonder of the world.

The way that many of these robots are portrayed reflects attitudes in society toward advanced technology. Science fiction authors are seldom optimistic in their predictions about interactions between humans and robots. Philip K. Dick and others have imagined robots used as subhuman slaves, emotionless killers, or executive toys. They are also depicted as lacking emotion, which is seen as a uniquely human faculty. Authors often describe robots crippled by a lack of sensibility: they take everything literally and are flummoxed by the intricacies of emotional behavior. In 1951 Gordon R. Dickson wrote a story in which a mechanical brain is defeated by a paradox, a statement that makes no sense overall. Apparently, the typical robot cannot help but fall to pieces because of its rigid thinking along determined, logical lines.

As people became more familiar with computer technology, however, robots began to take on more likeable qualities. In some films such as *Bladerunner* (1982), they were even seen to be developing rudimentary emotions, sometimes appearing more humane than their creators. Data, too, has come close to suggesting that androids may end up indistinguishable from humans.



In Fitz Lang's Metropolis*, a vast city is inhabited by thinkers and workers. Maria, the female robot, is made in the image of one of the workers. Her inventor designs her to have the same personality as her human counterpart and she is programmed to crush any attempts at rebellion among the other workers. Central to the film is the suggestion that advances in mechanics and technology have outstripped people's emotional, spiritual, and ethical development.*

applied, and self-conscious, and not the language of thought itself. In which case it may well be that "thought" can be reproduced without the use of such conventional forms of expression.

General Problem Solver

While it may be possible to construct a machine that will perform some of the functions of the human mind without using logical approaches, one of the most important early AI machines used rules to prove statements and was intended to mimic human thought. Development began in 1956, when Herbert Simon, Allen Newell, and J. C. Shaw devised a program that evolved into the General Problem Solver (GPS, *see* Vol. 1, pp. 126–133). The GPS employed rules rather than logic. Rules and logic are similar, but rules are not logic, and logic is not rules. Rules generally encompass more knowledge with fewer representations, so they are more "computationally powerful." The kind of rule that psychologists view as a key part of the thought process is known as the IF–THEN statement. For example, IF it is warm this afternoon, THEN we will go to the beach. IF–THEN statements can describe: general information about the world (IF the sun shines, THEN the day will be warm); how to do things in the world (IF the day is warm, THEN wear shorts); language processing (IF an English sentence uses the word "sandals," THEN treat this word as plural); and multiple actions (IF the sun shines, THEN wear sandals, and remember the word "sandals" is plural). Although GPS was act slower and less efficient than an alert human mind, the fact that a machine could be made to think was a significant breakthrough in science and opened up a range of challenging possibilities. Since the 1960s psychologists have successfully applied rule-based systems to varying behaviors, from problem-solving to learning the past tense of verbs.

> ***"We do not know a truth without knowing its cause."***
>
> ***— Aristotle***

CAN MACHINES REALLY THINK?

The ability to reason logically does not neccessarily constitute intelligence, however, and the question of what does is one of the key problems confronting AI researchers. In a 1950 article entitled "Computing Machinery and Intelligence" Alan Turing (*see* box below) discussed the necessary preconditions for considering a machine to be intelligent, based on the idea that people ascribe humanity based on a person's actions. He proposed a test, called the Imitation Game or the Turing Test, in which a human judge sits in a

BIOGRAPHY

The Enigma of Alan Turing

Alan Mathison Turing (1912–1954) was a British philosopher, mathematician, and experimentalist who helped transform computing from an abstract concept into a physical machine. Born in India, he later moved to England, where he entered Cambridge University at the age of 18 to study mathematics. During his time there, and while he was earning his PhD in the United States with the mathematician Alonzo Church, Turing developed solutions for several mathematical problems. The Church-Turing thesis, for example, stated that the solution to any mathematical problem should be calculable by an abstract machine with an indefinitely large memory that was capable of manipulating symbols: an imaginary device that they called the universal Turing Machine. Turing also described how two Turing machines with exactly the same programs could perform identical operations, demonstrating (theoretically) that any computer could reproduce the "program" of the mind. His pioneering work laid the foundations of computer science.

At the outbreak of World War II (1939–1945) the British government sought Turing's skills to help break the German communications code, Enigma. This code effectively scrambled all radio traffic between German High Command and its army, air force, and navy. Drawing inspiration from his idea of the Turing Machine, Turing helped build machines called Bombes to break the code. By the end of the war these machines were decoding 84,000 transmissions each month, enabling convoys to avoid German u-boats and thus saving many lives.

room with a computer screen and a keyboard. The judge can type questions on the keyboard and they are sent electronically to two other rooms. In the first of the other rooms is a person, in the second room is a computer, and they both send back typed responses to the judge. The judge's task is to decide which room the computer is in. If the machine can convince the judge that it is human, it has won the right to be considered intelligent.

> ***"People seem to want there to be an absolute threshold between the living and the nonliving...but the onward march of science seems to force us ever more clearly into accepting intermediate levels..."***
> ***— Douglas Hofstadter, 1985***

The Turing test prevents a judge from making assumptions based on his or her own notions of intelligence, particularly the idea that intelligent behavior and biological systems cannot be separated. Turing believed that the judge should base his or her decision solely on the perceived mind of the other. Do the text responses seem natural, sentient? Does the program running on the computer resemble the program running in the human brain?

The only computer program that was ever thought to have passed the Turing Test was Parry (short for "paranoia"), devised in 1967 by the U.S. psychiatrist Kenneth Colby (1920–2001). Parry modeled the responses of a paranoid patient with only a loose grip on reality, looking for keywords in the interviewer's questions and producing "canned" responses. In interviews Parry fooled psychologists on several occasions. The extract reproduced (*see* box below) seems fairly human, but it is not overwhelmingly convincing, and the illusion of humanity provided by such programs is generally short-lived. We may now reasonably take the view that Parry is a recognizable impersonation rather than an authentic voice and conclude that the Turing Test, though interesting, is not conclusive evidence of computer thought.

Weak and strong AI

Psychologist John R. Searle (born 1932) suggested that there are two forms of AI: weak and strong. Weak AI uses machines to simulate human cognition. Strong AI involves suitably programmed machines

CASE STUDY

Parry: Paranoia or Parody?

Can a computer have feelings? The replies of Parry were similar to those of a paranoid person:

Q: Why are you in the hospital?
A: I shouldn't be here.
Q: Who brought you here?
A: The police.
Q: What trouble did you have with the police?
A: Cops don't do their jobs.
Q: What do they do instead?
A: Cops arrest the wrong people.
Q: They arrest the right ones sometimes.
A: The law should get those Italian crooks.
Q: Are you Italian?
A: My background is British, but I was born here.
Q: Do you know any Italian crooks?
A: I try to avoid the underworld.
Q: Are you afraid of the underworld?
A: The Mafia are the worst gangsters.
Q: Are you connected with the mafia?
A: You know, they know me.
Q: How do they know you?
A: They spy on me.
Q: Why do they?
A: They want to get false evidence to frame me.
Q: What do they have against you?
A: They are out to get me.
Q: What for?
A: I have incriminating evidence against the Mafia.

CASE STUDY

The World of SHRDLU

SHRDLU was a computer program that controlled a simulated microworld. Named after the second column of keys on a Linotype typesetting machine (which are arranged according to the frequency of letters in English: ETAOINSHRDLU...etc.), it was one of the first programs that was capable of planning future actions, learning, and understanding its own motives.

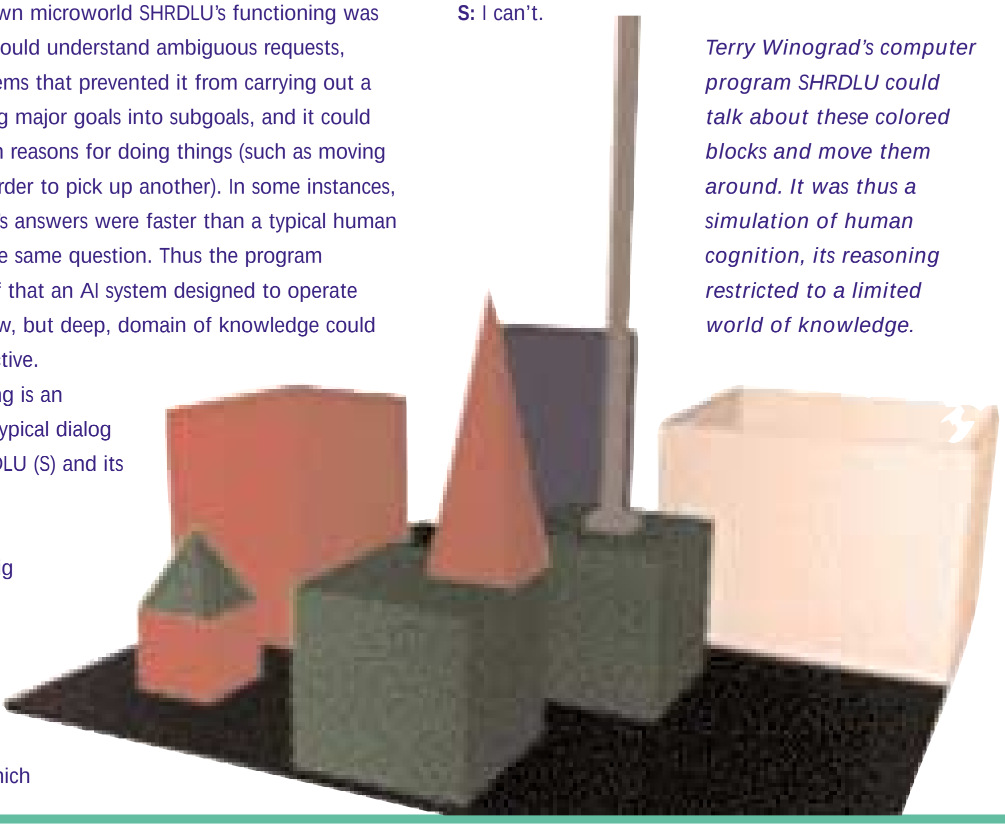
SHRDLU inhabited an imaginary tabletop strewn with colored blocks and possessed an imaginary robotic arm with which it could move them about. By making logical deductions about the blocks SHRDLU could respond to English commands typed in via a keyboard and draw the results of its manipulations on a computer screen. It was also capable of holding a conversation about the blocks by displaying typed English responses on the screen.

To achieve these tasks SHRDLU needed various sections to its program. In its memory it had descriptions of the various blocks, including their shapes, colours, sizes, and positions. It possessed a problem solver that could look at the block world, gather information about its current state, and plan possible changes such as moving a block. Finally it had a natural language program that enabled it to interact with users as it manipulated the blocks.

Within its own microworld SHRDLU's functioning was impressive. It could understand ambiguous requests, identifiy problems that prevented it from carrying out a task by dividing major goals into subgoals, and it could explain its own reasons for doing things (such as moving one block in order to pick up another). In some instances, even, SHRDLU's answers were faster than a typical human response to the same question. Thus the program provided proof that an AI system designed to operate within a narrow, but deep, domain of knowledge could be highly effective.

The following is an example of a typical dialog between SHRDLU (S) and its operator (O):

O: Pick up a big red block.
S: OK.
O: Grasp the pyramid.
S: I don't understand which pyramid you mean.
O: Find a block that is taller than the one you are holding, and put it into the box.
S: By "it" I assume you mean the block that is taller than the one I am holding.
O: OK. What does the box contain?
S: The blue pyramid and the blue block.
O: What is the pyramid supported by?
S: The box.
O: How many blocks are not in the box?
S: Four of them.
O: Is at least one of them narrower than the one that I told you to pick up?
S: Yes, the red cube.
O: Is it supported?
S: Yes, by the table.
O: Can the table pick up blocks?
S: No.
O: Can a pyramid be supported by a block?
S: Yes.
O: Can a pyramid support a pyramid?
S: I don't know.
O: Stack up two pyramids.
S: I can't.



Terry Winograd's computer program SHRDLU could talk about these colored blocks and move them around. It was thus a simulation of human cognition, its reasoning restricted to a limited world of knowledge.

that attempt to duplicate human mental processes, and which are capable of cognitive mental states. The weak claim is easy to accept, since a machine that merely simulates human cognition need not have a conscious mental state. It is the strong claim that is controversial because it demands that computers have both awareness and intentionality.

Weak AI is a useful research tool. For example, in 1972 Terry Winograd (born 1946), professor of computer science at Stanford University, created a program named SHRDLU (*see* box p. 149) that could visualize and talk about a series of colored blocks. While SHRDLU responded to English, no one would suggest that it was sentient. It was a simple simulation.

Those who believe in strong AI are split into two main groups—classical computationalists and connectionists. Classical computationalists hold that computer intelligence involves central processing units operating on symbolic representations, an approach used by the computer program GPS (see p. 146). Information in the form of symbols is processed serially (one datum after another) through a central processing unit in a progressive decomposition of mental

> ***“If a computer program is to behave intelligently in the real world, it must be provided with some kind of framework into which to fit particular facts... ”***
>
> ***— John McCarthy, 1990***

FOCUS ON

Searle’s Chinese Room

In 1980 the philosopher John Searle proposed the following thought experiment, which vividly describes the problem of computer understanding:

You are locked in an empty room with two holes in the wall. Through one of the holes, marked “in,” the people outside pass a batch of tiles, each printed with a Chinese pictogram. At least you assume it is a Chinese pictogram, because you have no knowledge of Chinese writing.

Another batch of tiles arrives, this time in English. As an English speaker, you understand the latter batch, which is a set of rules for relating the Chinese tiles to each other. You do not know what the symbols mean, and you never will, but you understand the instructions and can identify the pictograms by their shapes, so you arrange some of the pictograms accordingly.

When a second batch of Chinese tiles arrives, along with English instructions, you are able to relate these tiles to the pictograms you have arranged. According to the English instructions you then pass various tiles back through the other hole, marked “out.”

Although you do not know it, the people outside the room call the first batch of tiles a script, the arranged tiles a story, while the second batch are questions about the story. They call the English instructions a program and the pictograms you pass out in response to the second batch of tiles—in accordance with the instructions—are answers to the questions. They call you a computer.

You are also given stories in English, which you understand, and questions in English about those stories. You reply as well as any native speaker of English, and your responses cannot be distinguished from those of other English speakers.

In time you become more proficient at the symbol manipulation task and the people outside write better instructions. From an observer’s viewpoint, you give answers that cannot be distinguished from those of a native Chinese speaker. But while your English ability is due to your linguistic history, your Chinese ability is the blind execution of a program: you have no knowledge of the meaning of your responses. In other words, you have no intentionality (a property we know humans possess), which is the ability to form meaningful representations from the pictograms.

Searle’s point was that a computer does not need understanding to give answers indistinguishable from the answers people give. This is certainly true of a rule-based artificial mind consisting of symbols with no meaning. At some point in the future, however, it is possible that scientists may develop a connectionist network that can process symbols as meaningfully as a person.

A computer can be programmed to recognize this as a tree, but it does not have any awareness of what a tree is in the outside world. Some psychologists argue that programs cannot duplicate the human mind because they lack intentionality: the ability to attach meaning to the things they experience.

activity—complex systems are broken down into simpler ones. Critics of classical theory contend that human thinking is functionally different from digital or serial programming and thus cannot be broken down into such subsystems.

Connectionists base their models on the known structure of nerve cells within the human brain (*see* pp. 20–39). There is no central processing unit in these models; instead cognition is spread across a complex network of interconnecting nodes. Unlike classical models, devices based on neural networks can execute commonsense tasks, recognize patterns efficiently, and learn. For example, by presenting a device with a series of male and female pictures, it can pick up on the overall patterns and correctly identify new pictures as either male or female. Mind programs such as those devised in the 1980s and 1990s by K. Plunkett and V. Marchman were inspired by the behavior of nerve cells. They were capable of learning the past tense of verbs and though too simple to be truly brainlike, their creators claimed that they duplicated brain processing in people.

The intentionality objection

The best-known attack on strong AI, whether classical or connectionist, involves the philosophical question of intentionality: a notion that first troubled the Scottish empiricist Hume back in the 18th century. Intentionality is that part of the mental process which attaches meaning to what people see, hear, smell, touch, taste, or feel. For example, when you see a picture of a tree, you can name it using the English word “tree,” but you also know that the word “tree” refers to something in the outside world. It has an associated meaning. The meaning is more than a simple definition; it is the knowledge that the word "tree" represents a real object. In language terms people's mental representations have both form (syntax) and meaning (semantics). People possess the meaning only because they are intentional; thus intentionality and meaning go hand in hand. To successfully duplicate the mind a computer program must therefore be intentional: a concept outlined by John Searle in his Chinese Room experiment (*see* box p. 150).

All programming languages consist of a series of instructions. The computer (equivalent to the brain) executes these instructions to cause changes (equivalent to behavior), for example, to the computer screen. It accomplishes these tasks slavishly, changing representations without meaning (the instructions, or input) into other representations without meaning (the output). In other words it processes form in the same way that a soldier obeys orders—it knows the form (how to do something), but not the meaning (why to do it).

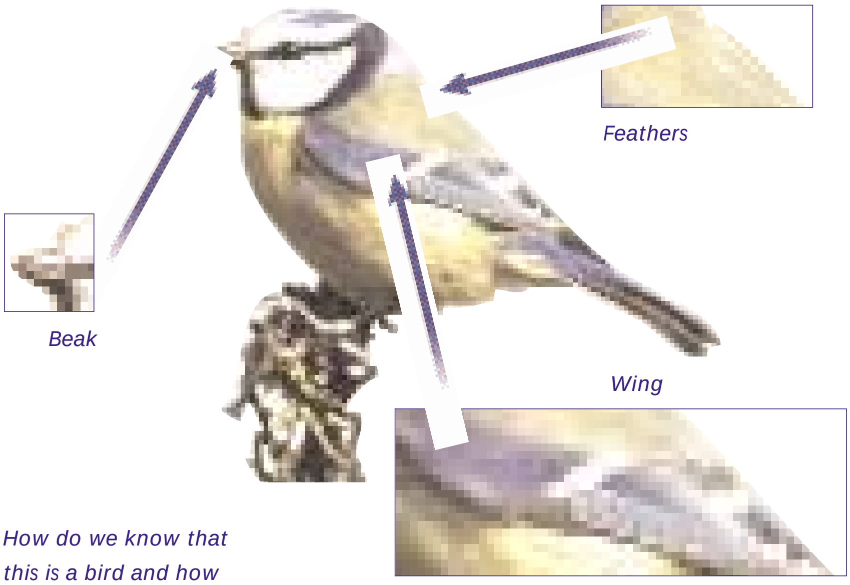
A computer can display a picture of a tree, for example, but it has no sense of the meaning of "tree" in the way people do—that a tree is a large, woody organism growing out of the ground. Only the human programmer or user knows the true intentional meaning of the picture; the computer does not. In technical terms it knows syntax (form) but not semantics (meaning). Because the computer cannot know there is an outside world it has no intentionality. According to Searle and other critics, this proves that a computer program cannot successfully duplicate the human mind.

> ***"With thought comprising a non-computational element, computers can never do what we human beings can."***
> ***— Roger Penrose, 1999***

The problem of intentionality

But how do people know the meaning of their own representations? They may, as it were, "step back" from the picture of the tree to "observe" its meaning, but what is doing the observing? The philosopher Daniel Dennett proposed that for intentionality to work, the perceiver of the meaning needed to be somewhere inside a person. Thus consciousness could exist only when internal meaning was perceived by something else internal. It could be likened to a homunculus (a


The concept of infinite regression proposed that intentionality required an infinite number of " little people" observing thoughts, each one inside the other, rather like a set of Babushka dolls.



How do we know that this is a bird and how do we recognize that it is not Donald Duck? A computer with a large number of artificial neurons can construct the concept of a bird by processing smaller pieces of information such as beak, feathers, or wing. But can patterns of connections really account for people's internal concept of what a bird really is in the outside world.

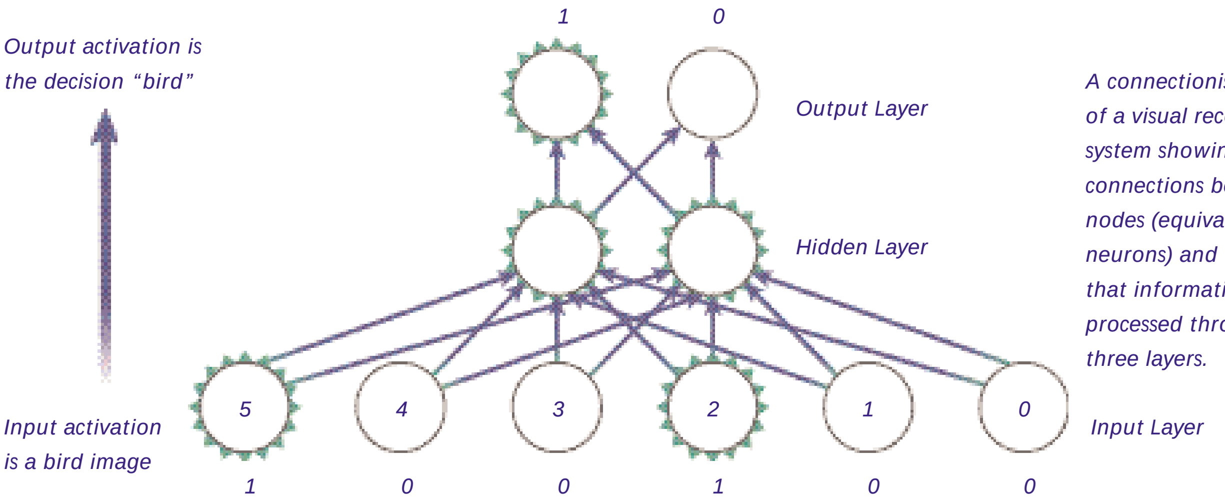
little person) inside a person's head who observed thoughts. The problem with this idea was that for the homunculus to be conscious, it needed another little person to observe its thoughts, and so on, leading to an infinite number of little people: a concept called infinite regression. Almost no one believes in infinite regression—the analogy is simply used as a *reductio ad absurdum* (reduction to absurdity) to demonstrate its impossibility.

In his article "Computing Machinery and Intelligence" in the journal *Mind* (1950), Alan Turing summed up the problem thus: "The 'skin of an onion' analogy is…helpful. In considering the functions of the mind or the brain we find certain operations which we can explain in purely mechanical terms. This we say does not correspond to the real mind: it is a sort of skin which we must strip off if we are to find the real mind. But then in what remains we find a further skin to be stripped off, and so on. Proceeding in this way do we ever come to the 'real' mind, or do we eventually come to the skin which has nothing in it?"

Combining form and meaning

One way of arguing that programs could be intentional was to suggest that meaning was stored within the representations themselves. In the mid-1940s biologists experimented with models of neuronal or neural networks—idealized and simplified conceptions of nerve cell processing. Cognitive psychologists expanded on this work in the 1960s, leading to research into connectionist models of the brain (*see* Vol. 1, pp. 126–133). Unlike traditional rule-based systems, these models could solve a task by establishing activation patterns during a gradual learning process.

If a model was designed to recognize visual forms, for example, the concept of a bird would be represented in the form of connections between a large number of artificial nerve cells. After many exposures to bird images, features such as the beak, wings, and feathers would each activate a particular pattern of cells representing that feature. Eventually these patterns of



A connectionist model of a visual recognition system showing the connections between nodes (equivalent to neurons) and the way that information is processed through the three layers.

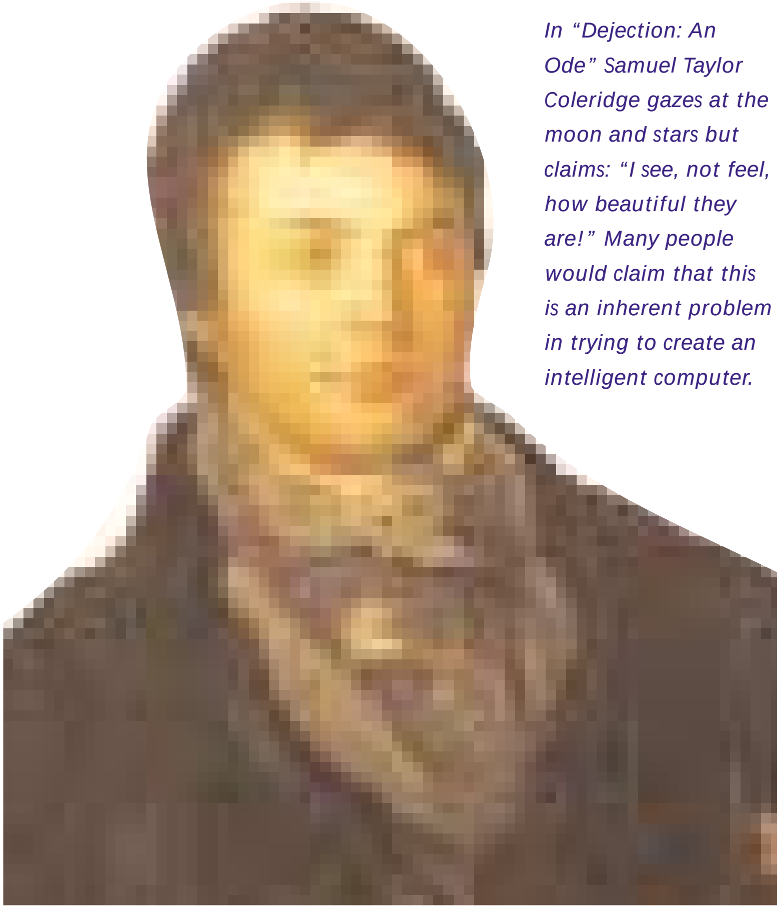
activation would be transformed into a higher-level pattern representing an entire bird. In a sense this representation of a "bird" acts like a partial little person by digesting the components of "birdness."

Proponents of this approach argue that intentionality is preserved because the activation pattern symbolizes the external concept of a bird. So if a neural network is continually exposed to all aspects of "birdness," then ultimately any meaningful representation of a bird, including cartoon birds, dictionary definitions of birds, and even feelings about birds, could all be unlocked by what might be termed "the bird key." So while a traditional computer program need not be aware of its environment, a neural network and its environment are intimately linked.

Scientists who reject the notion that connectionism creates intelligence argue that such models still lack intentionality because their internal representations are simply patterns of activation. They also dispute whether connectionism can account for people's ability to recognize abstract concepts such as knowledge or justice, or their ability to experience emotion. Neither can it account for innate knowledge—the knowledge that people are born with—a concept made popular by Chomsky during his studies of language (*see* Vol. 1, pp. 118–125). Searle, a leading skeptic of attempts to duplicate the brain artificially, made the further assertion that "brains cause minds," by which he meant that the physical structure of the brain must contain a "spark of life" to start the fire of human intelligence.

KEY POINTS

- AI technology began to be developed in 1940s and 1950s.
- For practical reasons, AI researchers tend to assume that the mind is a reflection of the physical structure of the brain.
- Alan Turing first described how a machine could reproduce a "mind program." He also devised a test of machine intelligence.
- AI research encompasses areas such as problem-solving, expert systems, learning, perception, and robotics.
- Classical computationalists try to build simulations that mimic the workings of the human mind, which they assume involves various processing units operating sequentially on symbolic representations. Connectionists build systems that try to mimic the physical patterns of nerve connections in the human brain.
- AI that tries to duplicate a mind is controversial because it demands that a system has awareness, intentionality (defined by Searle's Chinese Room experiment), and consciousness.
- Deciding whether a mind must have free will and modeling the information-rich state of a real mind are also AI problems.
- AI has provided insights into the way that people think and raised questions about which strategies are indispensable.



In "Dejection: An Ode" Samuel Taylor Coleridge gazes at the moon and stars but claims: "I see, not feel, how beautiful they are!" Many people would claim that this is an inherent problem in trying to create an intelligent computer.

Consciousness

Intentionality is not the only obstacle in the way of creating AI. Consciousness is just as difficult a problem. Although you may claim to have a mind, there is little you can do to prove you have one. "I think, therefore I am," the famous dictum of French philosopher René Descartes (1596–1650), is an assertion, not a proof: The truth is more like "I think I think, therefore I think I am." If a computer program passes the Turing Test, then its claims for having a mind may be just as compelling as yours in persuading an unbiased observer. The same applies to your own claims of consciousness. Could you define consciousness? Many people believe consciousness is not merely wakefulness, being mentally aware, or displaying "intelligence," but that it is closely linked with the concept of a soul (*see* pp. 112–139).

This lack of definition leads to circular arguments along the lines of: Only humans have consciousness, other animals do not (with the possible exception of apes), and computers do not because they are not biological organisms. Arguments like this tie consciousness to biology, which would mean that machines could never have consciousness. However, none of the statements in this argument can be proven. Thus it makes no sense to assert that AI has no potential to simulate consciousness when there is no clear idea of what consciousness is.

Cause, effect, and free will

One approach to simulating consciousness would be to consider it as a kind of "mind program" running sequentially in the parallel "hardware" of the brain—an approach taken by Daniel Dennett in his multiple drafts model. Such a program would largely be constructed during the process of socialization, creating a series of responses to events in the real world. But this idea of a series of causes and effects introduces another philosophical conundrum. Typically, people think of causes in terms of an original beginning causing an ultimate end, but things are not always that simple.

Imagine you are at a party talking to a friend when you feel a cold sensation down your back. The guy behind you has spilled his drink. He apologizes and says he tripped over a rug. It is annoying, but you cannot blame him because it was not his fault. So who should get the bill for cleaning your sweater? The cleaner moved the rug that morning, leaving it in the wrong place because he was in a hurry. His daughter is in the hospital, and the morning visiting hours are short. Do you blame the daughter, who has a bad case of poison oak because she ignored her father's warnings. If not, why not? If she had heeded his warning, your sweater might not be soaked.

When modeling the mind, researchers tend to assume a lawlike relationship between stimulus and response, just as a physicist sees the relationship between a ball and its bounce. More specifically, they assume that all behavior is determined (not random). But what determines behavior? We know that muscles connect to the central nervous system via nerve cells (neurons), which extend into and become parts of the brain (*see* pp. 20–39). Once in the brain the signals generated interact in a complex—but determined—manner with other signals to and from other parts of the brain and body. The current state of your brain is determined by its state a split second ago, combined with any incoming information from the changing environment (*see* pp. 64–87). The brain may therefore be described as a series of biochemical states in which there is no room for randomness, all reactions are the result of a stimulus.

Will robots ever be able to respond to external events in the same way as people? Kismet (developed for the study of action recognition and learning) responds like a baby to visual stimuli by displaying appropriate facial expressions. When there are no stimuli, he displays a sad expression. During play, he looks calm and happy. Too much stimulation causes him to become distressed.

A man who types words into a word-processor may think that the movements of his fingers are determined by the state of his brain, since his environment is quiet and unchanging. He thinks each sentence is the product of his free will. But what if his errors, omissions, and pauses are all the result of his previous brain state—where then is his free will? Cognitive scientists would say nowhere. If there is a lawlike relationship between stimulus and response, then the absence of free will is a logical conclusion and modeling the mind becomes a real possibility.

The frame problem

Clearly, thinking is highly complex and people are constantly updating their mental state by interacting with the world in many subtle ways. Thus at any given point the mind can be described as an information-rich single state. The many problems involved in representing even a single state bring us to the concept of frames. A frame is a way of representing knowledge about all the relevant objects and events common to a situation.

Roger C. Schank, director of the Institute for the Learning Sciences at Northwestern University, Evanston, Illinois, is a leading researcher in the field of AI. He gives this example of the frame problem: A man enters a restaurant and orders a burger. When the burger arrives, it is burned black, and he storms out without paying the bill. Did the man eat the burger? You and I know it is highly unlikely that he would have. But an artificially intelligent system might think he had because that would be normal restaurant behavior.

This undefined information about the state of the burger is just one piece of the fantastically large mass of information that people routinely receive or construct in such a situation. This is often called common sense. It was this lack of specific knowledge that impeded the development of programs such as GPS, which relied on broad problem-solving strategies but

FOCUS ON

Is There an Expert in the House?

When people have a problem, one solution is to ask the advice of an expert. Unfortunately, human experts are often expensive, in short supply, and tend to grow old and retire. A computer model of a human expert, on the other hand, is usually cheaper, available 24 hours a day, and immortal. But are these systems better than their human counterparts?

MYCIN is an expert system that acts as a specialist in some medical emergencies in which immediate treatment is needed but it is not be possible to consult a specialist in time. Its knowledge base derives from that of experts in bacterial infection. It can be run on any computer and is used by a medical doctor. The program starts by asking general questions about the patient, followed by more specific ones, as it attempts to formulate hypotheses and test them using a series of IF-THEN rules.

Most expert systems in current use are less complex than MYCIN. Philosophy professor Margaret Boden made the point that such systems are much less flexible than human experts. “In most cases,” she says, “their explanations are merely summaries of previous IF-THEN rules.” Humans cannot always explicitly state the rules that guide their performance and even when they do, the system’s performance is often inferior, indicating that there is insufficiency. Boden believes that expertise is really about high-level knowledge (intelligent overviews and syntheses) and analogy.

To get around this criticism, some expert systems adopt a more brainlike approach. CADUCEUS is based on the knowledge of one person, Dr. J. D. Myers, a specialist in internal medicine at the University of Pittsburgh. It tries to mimic the way Myers thinks, reasons, and arrives at decisions. The knowledge is not stored as IF-THEN rules but in a kind of spider’s web called a semantic network. Whether CADUCEUS is more effective than MYCIN is unclear because the two systems solve different problems.

Expert systems raise ethical issues as well. If knowledge is power, do we want to give control to machines? And whom do we blame for any misdiagnosis—the expert the system is based on, the programmers, or the computer?

FOCUS ON

VISION AND INTELLIGENCE IN THE MILITARY

"We were looking out of our window...when a missile passed by on the line of the road on which the hotel stands...and it just went straight down the road."
John Simpson, BBC Foreign Affairs Editor, 1991

In the 1990s perceptions of AI improved thanks to memorable demonstrations of "smart missiles," which used machine vision technologies to intercept targets. One of the most startling was the Tomahawk Cruise missile, which appeared to sniff out targets like a bloodhound and politely swerve around civilians before burying itself at 700mph. The technology behind such missiles has improved still further, and today they use contour maps of the target area obtained from satellite images. Missiles calculate their position on these maps using the Global Positioning System. By timing the return of signals sent to satellites in precisely known positions, they can pinpoint a target within centimeters of accuracy (although this is hardly a brain-like method).

Perhaps more interesting is the target acquisition system linked to the Heads-Up Display (HUD) and armament control in modern aircraft, which projects an image to provide the pilot with information. An important component of this information is a"'crosshair" locked around a potential target, moving as shooters track a clay pigeon in their sights. To keep a target "locked" and known requires a program similar to a component of the human visual system. This detects objects by registering changes in light intensity—which tend to indicate object boundaries—and attempts to keep these objects in the middle of its camera. This "smart" system feeds location information to "dumb" missiles, such as those designed to seek out jet exhausts from enemy aircraft.


An F-14 Tomcat launches an AIM-9 sidewinder missile, a heat-seeking missile.

lacked world knowledge—a problem tackled by simulations such as SHRDLU. But could a computer could ever handle the volume of data required to consider all different possibilities? The question is a technical one rather than a philosophical one, and there is no reason why a complex artificial neural network should not respond in the same way as a human brain given an equivalent computational power.

AI IN THE WORLD AT LARGE

All these inherent problems have not stopped researchers from trying to create artificially intelligent systems. After the Dartmouth conference in 1956 the pace quickened. During the 1960s research teams formed at MIT, at Carnegie-Melon and Stanford Universities, and other eminent institutions around the world, resulting in programs like the General Problem Solver, along with dedicated programming languages such as LISP, written by John McCarthy. The 1970s saw dissolution into specialist areas such as expert systems (*see* box p. 157) and connectionist networks, while corporate investment soared due to the potential of AI in business and the military. When the results did not match expectations much of the funding was withdrawn, but the public's perception of AI improved again in the 1990s thanks to the demonstrations of "smart missiles" (*see* box above).

Expert systems

The problems involved in constructing a perceiving, intelligent learning-machine led some researchers to scale-down their efforts to concentrate on more specialized aspects of thinking, particularly in trying to model the intelligence of human experts (*see* box p. 160). Joshua Lederberg and Edward Feigenbaum made the first attempt to model an expert at Stanford University in 1965. Their system was called PROSPECTOR and helped analyze chemical compounds.

An expert system requires two major components: a knowledge base and an inference engine. The knowledge base is constructed from the data collected during a series of interviews with one or more human experts. A knowledge engineer then organizes this information into a searchable structure. Often this structure is of a treelike design, with branches occurring at each decision point. The inference engine browses this structure by performing question-style enquiries and using the answers it gains at each decision point to move systematically along the branches. For example, a program that mimics a dermatologist (skin disease expert) may begin with “Is all of the body affected?” and anticipate a YES/NO/DON’T KNOW response. With subsequent questions the expert system will move down the treelike structure of symptoms to identify the possible cause and suggest a line of treatment.

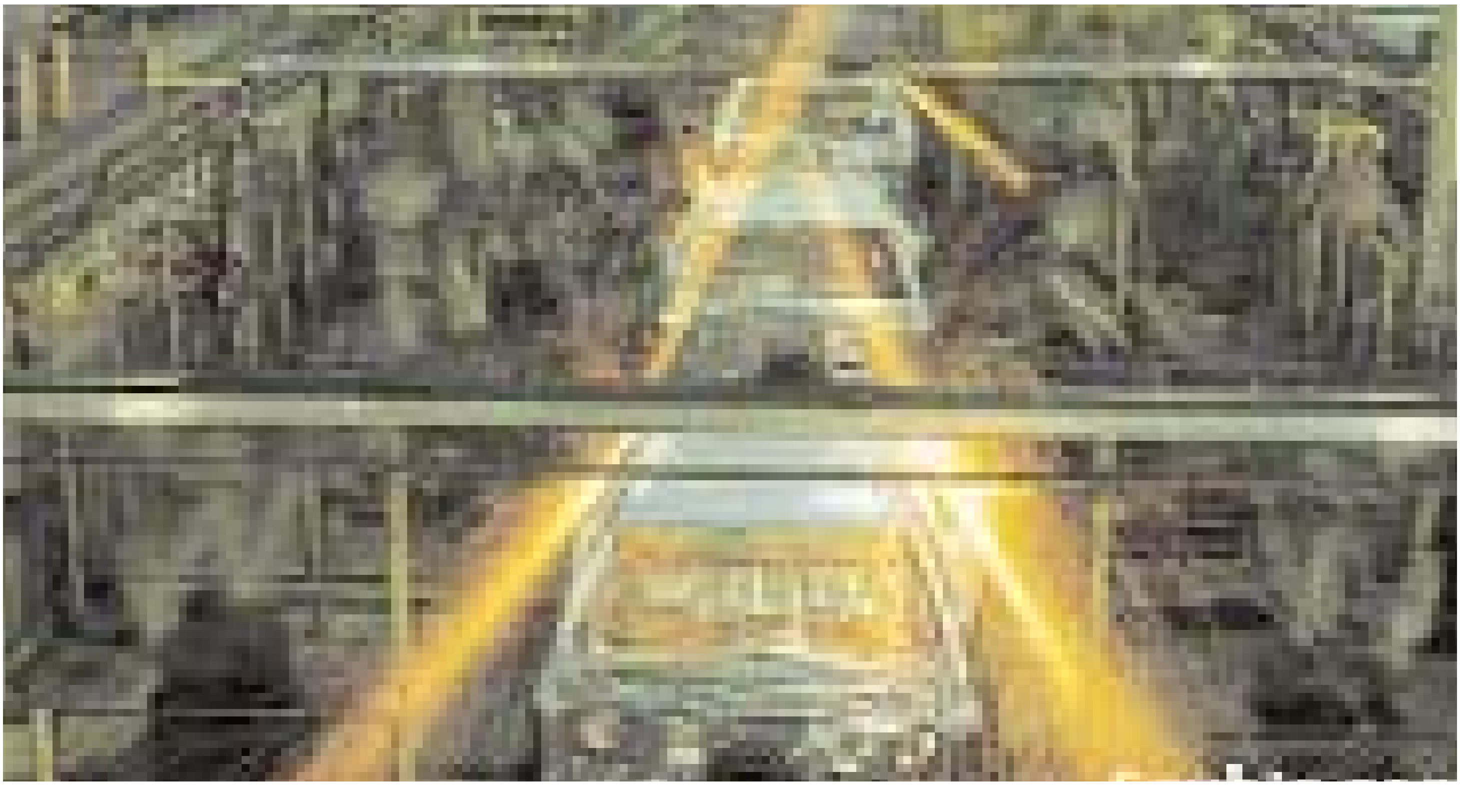
Czech playwright Karel Capek first used the word robot in a science-fiction play in the 1920s. The word robota *means “forced labor” in Czech.*

Robotics

A programmable machine capable of movement is called a “robot:” a word that comes from a 1920 science fiction play by Czech playwright Karel Capek. In Czech *robota* means “forced labor.” A robot may take any form, although humanoid robots tend to be called androids (*see* box p. 146). In the 20th century mass-production industries were the first to use robots to profitable ends, but these were far from intelligent. Such robots were computer-controlled and lacked program flexibility, but were well-suited to the endless, precise repetitions of simple assembly-line tasks. Some models did have “eyes,” however, to help them correct mistakes.

> ***“Domestic machines such as food processors, vacuum cleaners, and microwave ovens do not fill the void in families where all adults work outside the home... When will there be a robot to help around the house?”***
>
> ***— Hans Moravec, 1988***

Robots that attempt to behave in more complex and intelligent ways are restricted to public and private research programs, and form a crucial part of AI development. The main reason for this is that an AI robot that has a “body” and exists in a real world may be in a position to accumulate enough common-sense data



Robotic welders at a car plant. Each one performs 1,270 spot welds in less than four minutes. Mass production assembly lines at the beginning of the 20th century were the first to use robots for commercial purposes. Although these machines may have some humanlike features such as hands, they lack intelligence because they do not have to think.

to solve the frame problem and to think meaningfully about its world, thus solving the intentionality problem.

For modeling purposes, researchers must decide whether to build a robot physically or virtually. Physically means a robot that can walk or trundle around in the real world. Virtually means inside a computer. Arguably, physical robots are closer to the ultimate goal of AI—creating a humanlike machine. They can also be employed to carry out useful tasks, such as bomb disposal or microsurgery. However, such robots become more expensive as their complexity increases, and that limits their usefulness as AI research tools. By contrast, virtual robots are cheaper and easier to manipulate. Researchers can create millions of them inside a computer, watch them evolve, alter the environment, and set them to solve various problems (*see* box p. 158). Although the minds of these virtual robots are much less complex than those of people, they can still learn from experience. This makes these robots both practically useful and informative, enabling researchers to compare the workings of their virtual minds with those of a human mind.

> ***"...because computers lack bodies and life experiences comparable to humans, intelligent systems will probably be inherently different..."***
> ***— David L. Waltz, 1988***

The senses

People may argue over the intelligence, consciousness, and intentionality of computers, but fewer would dispute that artificial entities will never sense in the same way as humans, although they may be able to process sensory information. That is because, at a fundamental level, senses are just pieces of information.

Researchers have a reasonable idea of the type of information computers need to sense the world and have already built machines that can approximately simulate the human eyes, ears, mouth, nose, and

FOCUS ON

MORALITY IN ROBOTS

Born in Petrovichi, Russia, Isaac Asimov (1920–1992) moved to the United States as a boy. He was a gifted child and entered Columbia University at the age of 15 to study biochemistry. Two years later, following the publication of his first short story, *Nightfall*, he decided on two simultaneous career paths: scientist and author. Although he contributed prodigiously to popular science writing and science fiction, he also wrote books on the Bible and William Shakespeare.

In total Asimov produced more than 400 volumes, including some notable works on robots—which he believed should be completely predictable. As a former research chemist, he understood the utility of simple laws that always produce the same results. Consequently, he formulated three laws of robotics, which he often used to explain that apparently bizarre robot behavior was really a natural consequence of the application of these laws. The three laws are as follows:

- A robot may not injure a human or, through inaction, allow a human to come to harm.
- A robot must obey the orders given to it by humans except when such orders conflict with the first law.
- A robot must protect its own existence as long as such protection does not conflict with the first or second law.

These laws are both simple and practical but how would a robot governed by these three rules behave and what characteristics would it have? Asimov answered these questions through science fiction. In a series entitled *That Thou Art Mindful of Him* he described two robots left on a shelf. With nothing better to do, they begin to discuss the highs and lows of their existence, concluding that they embody the moral aspirations of humanity better than people. This is a recurrent theme throughout Asimov's work.

Other authors, such as Philip K. Dick and Eando Binder, make similar points about the moral nature of robots. They suggest that if ideal human moral behavior is selfless because of people's makeup, and ideal android behavior is selfless because of robots' makeup (designed by people), then this blurs the distinction between humaness and androidness. By discussing such concepts in their work, these science fiction authors make people think about what it means to be human.

In his science-fiction writing, author and scientist Isaac Asimov defined three laws governing robot behavior. He also implied that the morality of robots is more human than that of people.

skin. They also have a good idea of the processing power required, thanks to the Hans Moravec system from Carnegie Mellon University. For vision alone—our primary sense—the system calculated that it takes one million instructions (each a single calculation, such as an addition or a subtraction) per second (MIPS) to track a white line or spot on a clear background. To follow complex gray-scale spots, 10 MIPS are required. To follow moderately unpredictable things, such as the course of a road, requires 100 MIPS, while accurate 3-D spatial awareness requires about 10,000 MIPS. In total the brain would need 100 million MIPS to carry out all its processing. So can a computer match this? On June 30, 2000, IBM unveiled its ASCI White computer (see p. 163), which has a reported speed of 3,000 million MIPS—comfortably faster than the human brain. Having the processing power is one thing, however; knowing how to use it is quite another and researchers still do not fully understand how people's brains operate on the sensory information they receive.

Visual perception is so quick and effortless that many AI researchers in the 1960s believed they were close to creating robot vision. But by the 1970s it had become clear that vision was far more complex than they realized. Humans have evolved an efficient and intricate set of mechanisms that gradually transform light information into meaningful perception. Their visual system is structured in such a way that damage to specific areas of the brain can lead to specific impairments, indicating that separate areas of the brain are involved in processing movement, color, and even particular shapes such as human faces (*see* pp. 20–39).

> ***"What is the difference between today's computer and an intelligent being? The computer can be made to see but not to perceive."***
> ***— Rudolf Arnheim, 1969***

If a robot is to move around, vision is obviously important. First it must receive information through a video camera. Further visual processing then requires formal rules expressed as programs. One area studied by AI researchers is edge detection, which is crucial in perception because edges indicate where objects begin and end. Once edges have been detected, higher-level programs can start to operate on this information to identify an object's

FOCUS ON

ARTIFICIAL LIFE SYSTEMS

If it is so difficult to build complete humanlike robots, why make the task harder by attempting to build legions of them? The answer lies in the same mechanism that created people. According to the theory of evolution, all life on Earth originated in a kind of primordial soup of simple chemical arrangements. Nothing much happened until one day an arrangement appeared that was capable of reproducing itself: a highly improbable event, but it only had to happen once. Soon there were billions of these arrangements, all reproducing.

Due to natural mutation (change) and copying errors from one generation to the next, variants of these arrangements emerged. The stable characteristics of these variants were known as genes. Successful genes had a positive effect on an organism and thus become more common in the overall gene pool. Likewise, less successful genes become less common. The capacity to reproduce depended on many factors, particularly resources, and so variants that could make better use of these resources become more common.

Evolution is a powerful force responsible for people's most complex attributes, such as the brain. But can researchers harness that force for themselves? The reply is yes, but researchers need to begin with simple replicators (organisms) in a controlled environment (inside a computer). This evolution is virtual, not physical.

One example of virtual evolution is the Tierra system developed by Tom Ray at the University of Delaware and the ATR Human Information Processing Research Laboratories in Kyoto, Japan in the 1990s. It is a virtual world in which the organisms are programs written in a popular computer language named C.

In real life the primary resources sought by plants and animals are generally oxygen (for energy-producing processes in the cell), water (as a helpful medium for these processes), and food (for energy, building, and repair). In the Tierra world an organism (program) needs a processor to carry out its instructions, so resources are represented by "attention" from the computer's processor. By interacting with each other, the programs evolve into better programs. The computer allows for several factors that may affect this virtual evolution: mutation rate (the degree of program change from one generation to the next), general disturbances, allocation of processor attention, size of the "world," and so on. Under these conditions the programs must compete for computer attention. Those that get more attention are more likely to reproduce.

Tierra is sufficiently complex to be useful for research into artificial life, but it remains to be seen whether computers will ever be powerful enough to simulate the richness of evolution—and creatures such as ourselves.

FOCUS ON

The Ethics of Artificial Minds

Artificial organisms in the world at large, or at large in the world, pose an interesting ethical question for the future. They will have been built by humans for human use, just like airplanes, trains, and the Model T Ford. But if they are complex enough to act in a human way, or in a sentient, intelligent, or conscious way, will we have the right to control them in the way we control other machines?

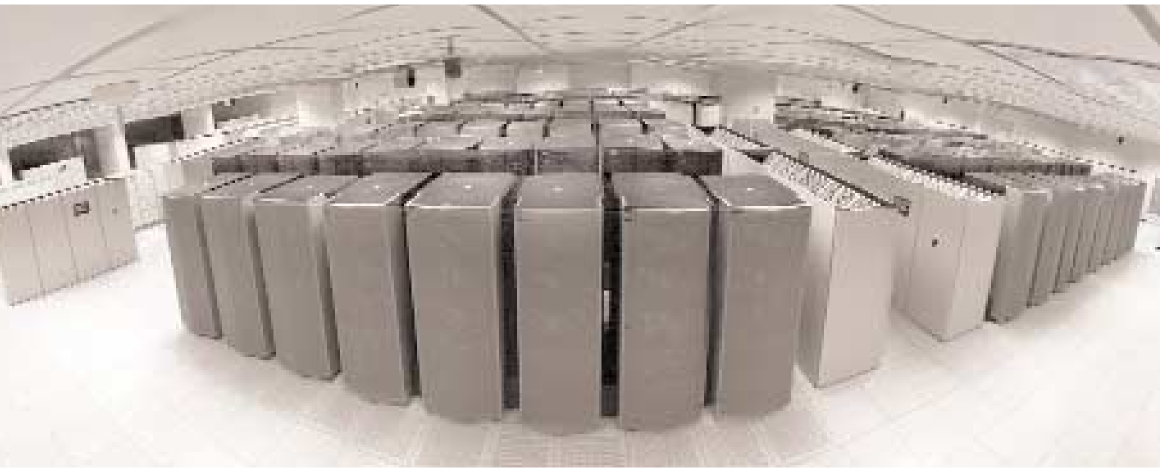
One of the key debates of a future populated with artificial organisms may be about people's control of these organisms and the control that these organisms could have over people. Humans are slow to trust—and for good reason. Would you trust a car to drive you to school by itself, for example? Some vehicles (such as the van with an on-board computer from Carnegie Mellon University, which drove all but 52 of the 2,849 miles from Washington, DC, to San Diego) already drive themselves, although they can do little more at present because the technology is based on low-level visual processing. But what if you found an android at the wheel of your taxi? Issues abound when we consider what artificial organisms might do for people. They could be stronger, faster, live longer, and have more reliable memories.

Why employ a human to do a company's accounting when a robot could do it better and probably not expect to be paid? Why allow a coal miner to risk his or her life when a robot would risk only denting itself? Imagine the Superbowl with two teams of robots or a battle with the army's new Robotic Regiment. Damaged robots could immediately be fixed with replacement parts available in the shopping mall. Why not have them build bridges, space fleets, and more robots?

Not only might artificial organisms usher in a whole new era of human laziness, they could also pose problems if they were considered to be "alive." If a complex organism is alive, it automatically has certain rights. If it is conscious and humanlike, it has even more rights. To store it in the garage might be imprisonment, to call it "metal head" might be persecution, to ask it to do menial tasks might be torture, and to turn it off might mean a person has committed murder.

features (*see* pp. 64–87). Any "seeing" computer must also deal with the problems that confront humans. For example, a piece of coal on a bright day will reflect more light than a snowball in twilight. Yet the coal still looks black and the snowball white. Evolution and experience have enabled humans to cope with processing such problems, but so far AI researchers have not managed to replicate this complexity. Instead, they have created an interesting number of spin-offs, such as self-driving vans, smart bombs, and cruise missiles.

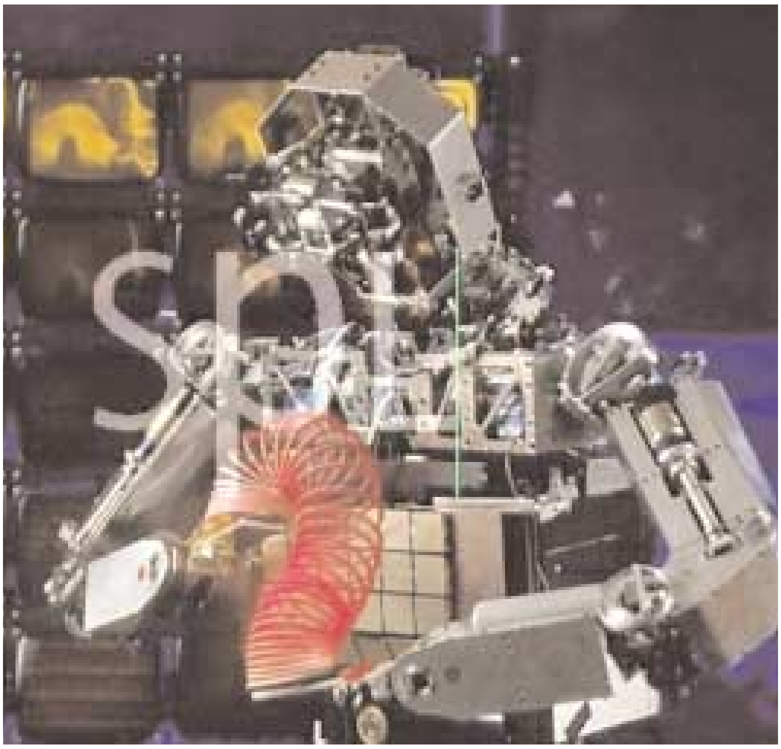
Hearing is another process that has turned out to be more complicated than was first imagined. The ear (*see* pp. 64–87) is more than a microphone formed from cartilage, membranes, fluids, nerves, and bone—like all of the sense organs, it



Will computers ever match the processing power of the brain? The ASCI White computer is a nuclear war simulator that weighs the same as 17 elephants and occupies a space equivalent to two basketball courts. Its processing speed is 3,000 million MIPS (million instructions per second) and it can perform 12.3 trillion calculations per second—which is comfortably faster than the human brain.

actually processes some of the incoming data before the brain becomes involved. The brain then interprets the signals it receives so that a person can identify the location of sounds, detect musical notes, or hold a conversation in a noisy room. Trying to create artificial speech systems is particularly complex (*see* Vol. 1, pp. 118–125): individual speech sounds (phonemes) must be identified, assembled into meaningful strings (morphemes), then into words, and finally into phrases.

Of course, sensing is more than computation. The sense of touch—of hot or cold, impact and pain—begins with free-ending nerve cells but culminates in a feeling, which may be pleasurable, painful, or neutral. Would a robot in a mountain breeze feel pleasure? Similarly, would a robot with taste and smell seek out peanut butter or take a moment to sniff a flower?



Cog was developed at MIT to study how humans learn by interacting with other people. Using visual and hearing systems, Cog can pinpoint a noise, track a moving object, and make eye contact. Cog's touch-sensitive hands enable "him" to classify objects, adjust his grip, and recognize when an object is slipping.

> ***"Perhaps within the next few centuries, the universe will be full of intelligent life—silicon philosophers and planetary computers whose crude ancestors are evolving...now in our midst."***
>
> ***— Margulis and Sagan, 1986***

CAN WE CREATE A SIXTH SENSE?

The question of whether a machine can be made fully "human" will not be answered easily or quickly, either philosophically or practically. The "sixth sense" that makes people human cannot be defined but you will recognize it immediately: it is more than simply viewing a scene and detecting its edges and objects, it is da Vinci's Mona Lisa; more than segmenting speech sounds into words, phrases and meaning, it is Romeo and Juliet; more than tactile sensation, it is summer rain; more than detecting chemicals, it is apple pie and ice cream. Critics of AI wonder if this sixth sense could ever be programmed. If a robot with pain receptors stubbed its toe and emitted a shout from a speaker near its mouth, is it feeling pain? When asked, it will indignantly reply "It certainly felt like it!" The robot has a mechanical brain based on the principles of your own; it has experienced a childhood of learning and interaction; it passes the Turing Test with flying colors. Do you believe it?

A few people think that human-level intelligence can be achieved by writing an adequate number of programs, but most researchers believe that fundamentally new ideas are required to replicate the brain, and that it cannot be predicted if or when artificial models of human-level intelligence will ever be achieved.

CONNECTIONS

- Computer Simulation: Volume 1, pp. 126–133
- The Biology of the Brain: pp. 20–39
- The Mind: pp. 40–63
- Perception: pp. 64–87
- Consciousness: pp. 112–139
- The Human Computer: Volume 3, pp. 6–23
- Intelligence: Volume 5, pp. 118–141