



### Artificial Intelligence: Current Status and Future Potential (1985)

Pages  
25

Size  
7 x 10

ISBN  
0309292719

Simon, Herbert Alexander; Naval Studies Board;  
National Research Council

 [Find Similar Titles](#)

 [More Information](#)

#### Visit the National Academies Press online and register for...

- ✓ Instant access to free PDF downloads of titles from the
  - NATIONAL ACADEMY OF SCIENCES
  - NATIONAL ACADEMY OF ENGINEERING
  - INSTITUTE OF MEDICINE
  - NATIONAL RESEARCH COUNCIL
- ✓ 10% off print titles
- ✓ Custom notification of new releases in your field of interest
- ✓ Special offers and discounts

Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

To request permission to reprint or otherwise distribute portions of this publication contact our Customer Service Department at 800-624-6242.

Copyright © National Academy of Sciences. All rights reserved.



The Charles H. Davis Lecture Series  
*Tenth Lecture*

**ARTIFICIAL INTELLIGENCE  
CURRENT STATUS AND  
FUTURE POTENTIAL**

*by*  
**Herbert A. Simon**  
Carnegie-Mellon University

Presented Before the Students and Faculty of  
The Naval Postgraduate School  
March 28, 1985  
*and*  
The Naval War College  
March 19, 1985

*Organized by*  
The Naval Studies Board  
of the  
National Research Council

*Sponsored by*  
The Office of Naval Research

**MAY 16 1989**

**PROPERTY OF  
NRC LIBRARY**

NATIONAL ACADEMY PRESS  
Washington, D.C. 1985

393

.456

10.10

21

# CONTENTS

The Charles H. Davis Lecture Series	5
Rear Admiral Charles H. Davis	7
Artificial Intelligence: Current Status and Future Potential	
Herbert A. Simon	11
A Quarter Century of Research	11
Defining Intelligence	12
Impediments to Growth	13
How Is AI Possible?	14
Testing and Applying the Hypothesis	14
Research on Chess	15
Expert Systems	17
The Moving Boundaries of AI	18
The Special Problems of Robotics	18
Systems for Scientific Discovery	20
Representation	21
Research on Learning	22
Conclusion	23
Curriculum Vitae, Dr. Herbert A. Simon	24
Search Committee for the Charles H. Davis Lecture Series	inside back cover



## THE CHARLES H. DAVIS LECTURE SERIES

**A**T THE CLOSE of that greatest of all contests of men and machines, World War II, Theodore von Karman could say, with deep personal conviction, that “. . . scientific results cannot be used efficiently by soldiers and sailors who have no understanding of them, and scientists cannot produce results useful for warfare without an understanding of the operations.” With such simple truths fresh on their minds, von Karman and his civilian and military colleagues proceeded to forge institutional links—such as the Office of Naval Research—through which they hoped to encourage an enduring partnership between the scientific and military communities. Though the intensity of the bond has fluctuated with the ebb and flow of international relations and internal affairs, the partnership has endured to produce a military capability but dimly perceived by those who established it. But the partnership is not self-sustaining; it requires the constant vigilance of those who have not forgotten the bitter lessons of the past, the outspoken dedication of those whose vision extends beyond the next procurement cycle, and, above all, it requires open communication between the partners. It is to this latter task that the Charles H. Davis Lecture Series is dedicated.

The lecture series is named in honor of Rear Admiral Charles Henry Davis (1807–1877) whose distinguished career as a naval officer and as a scientist so epitomizes the objectives of the series, and whose clear vision of the proper role of science in human affairs redounded to the betterment of all men. The topics and the speakers in the series are chosen by a Search Committee operating under the National Research Council of the National Academy of Sciences, and two lectures are presented each year before the students and faculty of both the Naval Postgraduate School in Monterey, California, and The Naval War College at Newport, Rhode Island. The series is sponsored by the Office of Naval Research.





Rear Admiral  
**Charles H. Davis**  
(1807–1877)

**C**HARLES HENRY DAVIS was born January 16, 1807, in Boston, Massachusetts. His education consisted of preparation at the Boston Latin School followed by two years at Harvard University (1821–1823). In 1823, Davis was appointed midshipman and sailed (1824) on the UNITED STATES to the West Coast of South America where he transferred to the DOLPHIN for a cruise of the Pacific. Returning to Harvard he continued to work on a degree in mathematics and is listed with the graduating class of 1825.

In 1829 Davis became passed midshipman and was ordered to the



ONTARIO (1829–1832) of the Mediterranean squadron. Later, while serving aboard the VINCENNES (1833–1835), he was promoted to lieutenant. Aboard the INDEPENDENCE (1837–1841) Davis made a cruise to Russia and then to Brazil. Throughout these early years at sea Davis continued to study mathematics, astronomy and hydrology. During this period one of his superiors would write of him, “C. H. Davis is devoted to the improvement of his mind; and his country may expect much from him.”

From 1842 to 1856 Davis undertook a number of special tasks and served on several commissions and boards. Notable among these was his participation in a survey of the New England coastal waters (1846–1849) during which he discovered several shoals that may have been responsible for a number of unexplained wrecks in the area. It was during this period in his career that Davis published “A Memoir upon the Geological Action of the Tidal and Other Currents of the Ocean” (1849) and “The Law of Deposit of the Flood Tide” (1852). He was also a prime mover in establishing the “America Ephemeris and Nautical Almanac” (1849) and supervising its publication at Cambridge, Massachusetts until 1855 and again from 1859 to 1862.

Promoted to commander in 1854, Davis resumed sea duty in command of the ST. MARYS in the Pacific (1856–1859). While he was captain of the ST. MARYS he was instrumental in securing the release of the adventurer William Walker and his followers who were besieged at Rivas, Nicaragua.

With the outbreak of the Civil War Davis was immediately appointed to a number of important positions. He became the executive head of the new Bureau of Detail for selecting and assigning officers. He was one of three officers appointed by Secretary Gideon Welles to the Ironclad Board which passed judgment on the plans and specifications for the MONITOR and other ironclads. Promoted to captain in November 1861, Davis participated in the development of plans for blockading the Atlantic Coast, planning the operation against Hatteras Inlet and Port Royal Channel, and the early naval strategy of the war.

During the operations against Port Royal, Davis served as captain of the fleet and Chief of Staff to Admiral Samuel F. Du Pont. He shares with Du Pont a great deal of the credit for the excellent plan of attack carried out on November 7, 1861. Later, as flag officer of the Mississippi Flotilla, Davis led successful engagements against the Confederate fleet which contributed to the abandonment of Fort Pillow and the surrender of Memphis. He was promoted to commodore in July 1862, and to rear admiral on February 7, 1863.

In late 1862 Davis returned to Washington to head the newly established Bureau of Navigation. From this position he worked closely

with such distinguished scientists as Joseph Henry and Alexander Bache to establish a “Permanent Commission” to advise the government on inventions and other scientific proposals which were being stimulated by the war. The Permanent Commission was established by the Secretary of the Navy on February 11, 1863 with Davis, Bache and Henry as members. However, Davis and his colleagues saw a wider need for cooperation between science and government and worked diligently for the establishment of the National Academy of Sciences. Their efforts were successful; President Abraham Lincoln signed a bill authorizing the establishment of the Academy on March 3, 1863.

In 1865, Admiral Davis was appointed superintendent of the Naval Observatory in Washington. In 1867 he returned to sea in command of the South Atlantic Squadron. Back in Washington in 1869 he was made a member of the Lighthouse Board and commander of the Norfolk Navy Yard. He later resumed his post as superintendent of the Naval Observatory where he served until his death on February 18, 1877.



**DR. HERBERT A. SIMON**  
Carnegie-Mellon University

# ARTIFICIAL INTELLIGENCE CURRENT STATUS AND FUTURE POTENTIAL

HERBERT A. SIMON  
Carnegie-Mellon University

**A**T THE OUTSET, I need to clarify what I mean by the term *intelligence*. In a military context, intelligence can have two quite distinct meanings: first, it can refer to information of significance to military operations and to the means for securing and analyzing it; second, it can refer to the faculty of the human mind and brain that enables us to think and learn. It is the second meaning that was intended by the inventors of the label *artificial intelligence*, and that will be my meaning in my talk today.

I might prefer a different label. In our early work, Al Newell, Cliff Shaw, and I referred to what we were doing as “complex information processing.” But since that’s a rather bland phrase compared with the challenging claim that our field has something to do with intelligence, it is the AI label that has stuck.

## A QUARTER CENTURY OF RESEARCH

AI has been very much in the public eye during the past two or three years. Since almost all weekly and monthly journals and many newspapers have done feature stories on it, I am sure that it has not escaped your notice. Some of you may be surprised, however, to learn that it is not at all a new thing. Research on artificial intelligence began in 1955 with a computer program called LT (Logic Theorist), which was capable of seeking and finding proofs of theorems by a process of heuristic, or selective, search. By 1958, computers were designing motors, transformers, and generators automatically in companies like Westinghouse, and a paper was published by Goodwin in that year titled “Digital computers tap out designs for large motors . . . fast.” In 1960, Clarkson showed how a computer could choose stocks and bonds for an investment portfolio in much the same manner as that task was done by a bank investment officer.

## DEFINING INTELLIGENCE

Why do I cite these programs, now a quarter century old, as examples of AI? To answer that, we have to agree on what intelligence is. We judge our fellow humans—and ourselves—to be intelligent when we can perform certain kinds of tasks: solving problems, making decisions, learning. Of course we also sometimes judge a person to be intelligent if we see a thoughtful frown, but that isn't always a reliable cue. It is safer to use some kind of intellectual task, perhaps of the kinds that are found on intelligence tests.

If a human being did any of the tasks I mentioned earlier—proving a theorem, designing a motor, choosing a sensible trust portfolio—we would probably agree that intelligence had been exhibited. Similarly, we can say that artificial intelligence has been exhibited by a computer when it has done something that would have required intelligence in a man or woman. Artificial intelligence is the study of computers doing intelligent things—things that would require intelligence of people.

The possibilities of AI have also long been of interest to military organizations. I remind you of the evolution of early warning systems for aircraft, which have gradually been automated over the past 30 years by incorporating more and more intelligence into the computer components of the systems. In the same way, more and more of the intelligence required for landing planes has been taken over by automatic systems. There are many other military applications, existing and planned, that illustrate the automation of intelligence. We are so accustomed to thinking of these as “automation” that we seldom consider the level of intelligence that would be required of a human being performing the same task. But there is no sharp line between automation and artificial intelligence: the one gradually merges into the other with the introduction of new and more powerful techniques of automation.

We tend to think of a computer application as automation if the task performed has a well-defined goal and structure and well-defined alternatives, and especially if it is performed in a carefully designed, simplified environment. Thus, automatic welding in a factory setting is automation and is not usually regarded as artificial intelligence; welding performed on pipes on an irregular sea bottom may require a device that is much more flexible in its responses, hence one that genuinely incorporates techniques of artificial intelligence. Automation deals with the domain of well-structured problems, AI with the domain of ill-structured problems, and there is no sharp boundary between the two domains.

## IMPEDIMENTS TO GROWTH

If AI is already 30 years old, why are we just beginning to learn about it now? Why can we point to only a few real-world applications, and these mostly quite recent? Until the past few years, the growth of the field has certainly not been precipitous. What have been the limits on its development?

I think there is no single explanation for the slow growth of applied AI, but several contributing factors. First, it has been paced by limits on the processing speed and memory size of computers. The expert systems that we are beginning to hear of today are large programs, with large associated data bases. Until megabyte memories were common and cheap, it was not realistic to think of large programs of this kind.

Computer speed and size, and computer architecture, did not much impede the *conceptual* growth of the field—that has been limited mainly by the imaginations of the researchers. Computer hardware probably did limit the rate at which basic research ideas could be translated into practical application. The principle of investment in securities by computers could be demonstrated by Clarkson in 1960; however, there was no possibility, at that time, of building a system suitable for practical everyday use.

Today we hear a great deal about prospective new computer architectures, so-called supercomputers and parallel computers. While it is always nice to have larger and faster machines, I do not think these developments are critical for the conceptual development of AI, except possibly in the domain of visual and auditory pattern recognition. Nor do we have to wait for new programming languages to advance AI. The principal programming tools have been available for a long time: list processing languages from the beginnings of AI, and so-called production systems for the past decade or more.

One other impediment to the rapid growth of AI should be mentioned: skepticism. People, including some members of the computer science community, have often been extremely skeptical that computers could exhibit anything that could reasonably be called intelligence or could compete with humans in the quality of their performance of professional-level tasks. Only concrete demonstration with running and debugged computer programs has been able, step by step, to gain ground against that skepticism. The boundary is redrawn each year as computers demonstrate their competence in new domains, but the skeptics simply retreat to the territory that has not yet been explored.

Skepticism has impeded progress in a number of ways. It has made it more difficult to fund research in AI. It has made it difficult, until

quite recently, to attract large numbers of the ablest graduate students to the field. And it has diverted the energy of AI researchers themselves to debating with the skeptics about the possibility of AI, instead of simply writing programs that exhibit it.

But I am getting ahead of my story. I should say something about why the skepticism is unwarranted—about how artificial intelligence is possible.

## HOW IS AI POSSIBLE?

Artificial intelligence research rests on a hypothesis, sometimes called the physical symbol system hypothesis, which states:

The necessary and sufficient condition for any system, biological or mechanical, to be capable of thought and intelligence is that the system be a physical symbol system: that is, that it be able to input (*read*) symbols, output (*write*) symbols, *create structures* of symbols related in various ways, *store* symbols and symbol structures in a memory, *compare* symbol structures for identity or difference, and *branch* (adapt its behavior) on the basis of the outcomes of such comparisons.

By a symbol is meant any pattern, whether built of ink, of chalk, of electricity and magnetism, or of neurons, that can be compared with other patterns. The electromagnetic fields stored in the memory of a computer are patterns, hence symbols. They may be used to represent numbers, words, or even pictures and diagrams. So the physical symbol system hypothesis claims that any system that can read, write, relate, store, and compare patterns and branch can be programmed to behave intelligently.

The hypothesis has two immediate corollaries. If it is correct, then it follows that:

1. Computers can be programmed to think (since they obviously possess the symbol-processing capabilities listed above).
2. Humans use symbolic processes, implemented by the nervous system, to think.

### TESTING AND APPLYING THE HYPOTHESIS

There is no need either to accept or to reject the physical symbol system hypothesis on faith. It is an empirical hypothesis to be accepted

or rejected, like any other scientific theory, on the basis of factual evidence. Much of the research activity in AI, and in the allied domain called cognitive science, is aimed at performing such tests.

To test the hypothesis that computers can be programmed to think, we construct computer programs to handle a variety of tasks of the kinds that require thinking in humans. To test the hypothesis that a thinking human being is a physical symbol system, we try to write computer programs that think in as humanoid a fashion as possible (taking the same steps, making the same mistakes as people) and then compare the behavior of these programs, in detail, with the behavior of people performing the same tasks.

From this description, you can see that research in AI has two distinct goals, which are complementary but directed at quite different applications. First is the goal of supplementing human brain power with intelligent machine power, of augmenting the amount of intelligence available for dealing with human problems and decisions. The first stage of the Industrial Revolution was concerned with supplementing human muscle with the power of machines; this second stage of the Industrial Revolution is concerned with supplementing human brains with the intelligence of machines. Artificial intelligence thereby becomes a major direction in which we seek to increase productivity in our society, so that we will have the resources to meet the human, social, and security needs of that society.

The second goal of AI, or of the related field called cognitive science, is to understand and improve human thinking and learning processes, using the computer as a central tool to build models of those processes. This goal is a part of the task of experimental and theoretical psychology. As understanding is gained of human thinking and learning, it can be applied to improving our decision-making processes in organizations and our teaching and learning processes in the schools.

#### RESEARCH ON CHESS

Turning from a general characterization of the field, I should like now to illustrate my thesis with examples, the first being research on the game of chess. Chess may appear to be a trivial domain for serious research, but we should recall that some of the most important advances in genetics in this century were made using fruit flies as experimental organisms. Fruit flies are of only slight economic importance, and even less esthetic appeal. But they offered certain advantages (rapid breeding, simple care, large chromosomes) for experimentation, and a large body of knowledge accumulated about this domain.



Chess has also proved a valuable experimental domain, this time for the accumulation of knowledge about intelligence. The game has little regularity—few mathematical theorems can be proved about it. It is complex enough to engage bright people for a lifetime. It even has characteristics that are relevant to military competition. So chess has become the fruit fly of artificial intelligence.

The first thing that was learned about the game is that it is impossible to find good moves by solving an optimization problem. The space is simply too large for computers present or prospective to explore systematically: some  $10^{120}$  possible games. It is a good environment in which to learn how people deal with complexity that forbids exhaustive search or exact algorithms.

A study of human chess experts—grand masters—began to reveal the bases of their skill. The grand master, it turns out, is able to recognize any one of a vast number of features when that feature appears as a pattern on a chessboard during a game. A grand master's repertoire is at least 50,000 such patterns, and with each pattern is stored in memory information about what to do when the pattern is seen—for example, "If you see an open file, put a rook on it." Now it is not true that every time a grand master sees an open file he moves a rook to it. But it is true that he notices the features and thinks of moving the rook. If he did not think of it, in a matter of seconds, he would not be a grand master.

A large part of the grand master's skill, then, is the skill of recognition, those 50,000 patterns or "chunks" as we call them in psychology. It is this recognition ability that allows him to play 50 games simultaneously with weaker opponents, at 10 seconds a move, waiting until the opponent makes an error that is recognizable as one of these features and exploiting the error in a standard fashion.

The same knowledge exhibits itself in other ways. Display a chess position to a grand master from some well-played game that is unknown to him. After 10 seconds remove it and ask him to reconstruct the position. He will replace correctly 23 or more of the 25 pieces on the board. A weaker player, even a good amateur, faced with the same task will replace only 6 or 7 correctly. Must you have some kind of special visual imagery to be a chess master?

Put the same 25 pieces on the board but arrange them at random. Again, the amateur will succeed in replacing about 6—and so will the grand master! What is involved is not visual imagery but the 50,000 chunks. The game position is made up of familiar configurations, the random position is not. The same prowess can be demonstrated in expert bridge players, poker players, or devotees of any activity who have put in 10 years learning their trade.

## EXPERT SYSTEMS

That brings us to the topic of expert systems, the principal form that applied AI takes these days. An expert system is any system that can perform professional tasks in some domain at an expert level. The program may achieve its expert performance by borrowing some of the tricks that human professionals use in performing the same task, or it may rely largely on the brute force speed and memory size of modern computers. Most existing expert systems combine these two sources of power: they use heuristics, or rules of thumb, borrowed from people, in order to search selectively and intelligently, rather than depending on massive search by pure trial and error. However, they are usually prepared to carry out more extensive searches than humans can, to compensate for possibly incomplete knowledge and heuristics.

For example, the very strongest current chess playing programs can perform at master level. They may examine as many as a million, or even several million, branches of the game tree before they make a move. Of course a search of  $10^6$  possibilities is still very selective when compared with the total search space of  $10^{120}$  possibilities. Nevertheless, human masters and grand masters do as well or better, although they seldom examine more than about 100 branches of the game tree. The chess programs are still trading off computer speed against their limits of chess knowledge.

In the same way, medical diagnosis systems, ore prospecting systems, and other expert systems that have been developed search selectively, but not as selectively as human professionals, and compensate for their lack of selectivity with their speed.

Described in general terms, most expert systems have a common architecture, consisting of a large data base, an "indexing" system for accessing the data base on the basis of the cues presented in the problem situation, and capabilities for making inferences from the knowledge they draw from the data base.

A medical diagnosis system like CADUCEUS or MYCIN, for example, contains a large store of medical information, consisting of disease entities (possible diagnoses) associated with symptoms (cues). Given an initial set of symptoms, the system recognizes one or more disease entities that are indexed by these symptoms. On the basis of its knowledge, it then requests additional information and tests to allow it to discriminate among the alternative diagnoses. It may also draw inferences from symptoms in one part of the body to originating causes in other parts of the body. As evidence accumulates, it is able gradually to assign a much larger weight to the probability of one diagnosis than to the others and to arrive at a final decision.

An expert program in chemistry for synthesizing organic reactions would be organized in a similar way, but it might make greater use of means-ends analysis in solving its problems. Given a set of initial reagents, their costs and their chemical reactivities, the system is asked to design a set of reactions that will produce a desired substance. Starting with the desired molecule, the system reasons backward to simpler substances from which it could be synthesized, and from these to substances available among the reagents. In choosing a desirable reaction path, it can take into account not only the characteristics of the chemical reactions, but also the costs of ingredients and energy.

I cannot give you a list of the expert systems that have been built and used up to the present time, both because the list is quite long and because new systems are being created almost daily. Neither can I claim a comprehensive knowledge of them or of the extent to which they have actually been applied to real-world situations. In addition to the examples I have already mentioned, there are systems for interpreting the data in mass spectrograms, systems for configuring complex computer systems to meet the requirements of particular customers, systems for scheduling large job shops, systems for interpreting oil well drilling logs, and others. And as I mentioned earlier, the boundary is very vague between the expert systems that we call artificial intelligence systems and other kinds of highly automatic control systems.

## THE MOVING BOUNDARIES OF AI

The research strategy of artificial intelligence has followed the usual path of starting with the simplest kinds of situations—highly structured “toy” problems—and, when these have been mastered, moving on to situations that are more complex and less well structured. The field advances gradually from the well-structured to the ill-structured, and from data-poor to data-rich domains. That kind of progress is continuing, and it is useful to ask what lies ahead.

### THE SPECIAL PROBLEMS OF ROBOTICS

Robotics is that special branch of artificial intelligence that is concerned with systems that perform physical actions on their environments. What is the difference between robotics and the forms of factory automation with which we have been familiar for many years?

In factory automation, the focus is placed on simplifying tasks

sufficiently and standardizing the environments in which they are performed, so that they could be done by machines without human intervention. A typical example might be the machines for boring cylinders in automobile engine blocks. The underlying analytic techniques for designing and building such systems have been drawn mainly from control theory.

Current robotics research aims at relaxing the limits on automation in two related directions: first by providing the machines with sophisticated sensory organs—"eyes" and "ears"—so that they can locate and manipulate the materials they are dealing with, and flexible manipulative faculties—"hands" and "legs"—so that they can perform complex actions in continually varying circumstances.

The human eye and ear and limbs evolved along with the older parts of the brain that we share with our mammalian cousins. Nature has had 400 million years to shape and perfect mammalian (including human) sensory and motor systems. They are very sophisticated and complex systems, capable (e.g., in the retina of the eye) of rapid parallel computation, and highly attuned to their tasks.

By contrast, the new parts of the brain, which we are especially proud of because they distinguish us from other species of animals, have been evolving for only a couple of million years. The parts of the brain that we use for abstract, professional thinking are still rather simple and crude. The new brain is a slow, serial, one-thing-at-a-time system. There has not yet been time for fine tuning.

It should not have been a surprise, therefore (although in fact it was), that progress in artificial intelligence research was easier and more rapid in imitating the functions of the new brain than it was in imitating sensory and motor functions. The bottleneck in robotics research today, where progress is being made but relatively slowly, is in the design of sensory and motor organs, not in the design of systems capable of reasoning about sensory information and turning it into plans of action. It is easier to automate professors and naval planners than it is to automate bulldozer drivers.

The reason, of course, why we want our new robots to have sophisticated sensory and motor capabilities is that we want them to be capable of operating in natural environments that have not been presimplified to facilitate automation. The problems are different both from those of traditional factory and office automation, in which the complexity of the environments could be controlled and reduced, and from planning tasks, in which the environment can be represented in a smoothed and simplified model. A robot vehicle, operating in a real environment (an unmanned and untethered submersible, for example), must be able to cope with the unexpected and the variable. It must be

able to readjust its picture of the external world continually as its sensors provide it with new information.

These are some of the reasons why robotics is perhaps the most challenging—and sometimes frustrating—branch of artificial intelligence today, and why we should expect steady progress but not sudden miracles of accomplishment.

#### SYSTEMS FOR SCIENTIFIC DISCOVERY

To take a still deeper look into the future, I would like now to describe work that is still in the basic research stage and far from practical application. I refer to computer systems that, by examining empirical data, can discover new scientific laws. It is sometimes thought to be impossible for a computer to discover anything new, to do anything creative. After all, a computer can do only what it is programmed to do. But that perfectly true statement does not mean that a computer always does what its programmer thought it was going to do. Nor does it mean that a computer cannot be programmed to do things that its programmer would be unable to do. Nor does it mean that it cannot be programmed to do very general things: specifically, to explore a body of data looking for regularity (just as it might explore a chess position looking for an unsuspected pattern of moves).

An example of a discovery system is BACON, a program developed by Patrick Langley, Gary Bradshaw, Jan Zytkow, and myself. BACON's expertise derives from a small set of rules of thumb that guide its searches through data. It looks for correlations among variables, it looks for invariant relations, it invents and introduces new concepts that may simplify the relations it has found, it looks for symmetries among variables and possible laws of conservation, and it is alert for integral ratios among variables of which all are integral multiples of some single number.

When given the distances of the planets from the sun and their periods of revolution, BACON discovers (in less than a minute) that the periods vary as the  $3/2$  powers of the distances. That is Kepler's Third Law, an important discovery of the seventeenth century. When given data on the accelerations of blocks of different masses connected by a spring, BACON discovers the law of conservation of momentum and, in the process of doing so, invents and applies the concept of inertial mass. Up to the present time, we cannot claim that BACON has discovered anything truly new, because we have exercised it on historical data in order to calibrate it against major scientific discoveries. Of

course, BACON has no reason to know that it is merely rediscovering laws that were already in the textbooks—and, because it has no way of knowing that, its rediscoveries are no less difficult and innovative than the original discoveries.

BACON is not the only discovery system that has been built. There are also the AM and EURISKO systems constructed by Douglas Lenat, which have labored mainly in the domain of mathematical discovery, and the META-DENDRAL system that discovered certain chemical laws relating to mass spectrogram analysis. Only the future will tell what potential these systems have for practical application. But they have already demonstrated conclusively that computers can be programmed to explore and invent.

## REPRESENTATION

In human problem solving, success often depends on representing the problem correctly. There are many problems that are hard to reason about in words, but whose solution becomes easy to find when the proper diagram is drawn. An important task in basic AI research today is to discover why this is so—how diagrams are used in the process of solving problems.

For example, I ask you to imagine a rectangle that is twice as wide as it is high. Now draw a diagonal from the NW to the SE corner, and a vertical line from the middle of the top to the middle of the bottom of the rectangle. Do the two lines intersect? Of course. How do you know? You can “see” it in your mind’s eye. Now proving formally that those two lines intersect is a formidable mathematical task, requiring either analytic geometry or topology. Do you think you could construct such a proof? Yet that is, in effect, what your mind’s eye did in just an instant. It is a powerful inference engine.

A large fraction of the resources of our society are spent in educating ourselves (you must count the students’ time as well as the teachers’). Insight into mental processes—those used, for example, in visual imagery—can lead us to greatly improved techniques for teaching and learning, because in AI there are two complementary routes to enhancing human intelligence. One is to build intelligent machines that can augment it. The other is to improve our skills in using our own minds for learning, thinking, solving problems, and making decisions. The latter route, the enhancement of human performance, may in fact be the more important one.

## RESEARCH ON LEARNING

This brings me to my final example of ongoing research in AI: research on learning. I have several times mentioned production systems, which are simply computer programs that use a particular kind of programming language. Production systems are attractive for AI applications because they are easily modified simply by deleting instructions (productions) or inserting new ones. This simplicity of structure suggests that they might be programmed to modify themselves in adaptive ways—to operate as *adaptive* production systems. This has now been accomplished in a number of laboratory applications.

The adaptive production system is exposed to a worked-out example of a problem solution—say the three or four steps that are involved in solving a simple linear equation in one variable. The system examines the example, comparing successive steps to see what changes have been made and what has motivated these changes. The reasons for the changes are detected by determining what progress the step contributed toward the final result. The program then constructs new instructions that, given the same cues, will take the same actions. These new instructions, when assembled and inserted (automatically) in the program, constitute an effective program for solving equations in algebra.

On the basis of experience in constructing and testing adaptive production systems of this kind, we can design learning experiences for human students that will allow them to acquire new skills (to build new productions in their minds). About a year ago, this idea was used to test whether secondary school students could learn to factor quadratics, simply by being exposed to a carefully planned sequence of worked-out examples. The experiment, which was carried out in a number of schools in the People's Republic of China, was highly successful. Almost all the students learned to factor quadratics in less than 20 minutes. Similar experiments are now being carried out in the United States—and of course we will need much more experience with these ideas before we will know how to use them effectively to improve the teaching of mathematics.

Teaching and learning, as we do them today, are highly pragmatic activities, shaped by practical experience but with little underlying theory. Artificial intelligence has now reached the point at which it can tackle tasks at the level of complexity of school and university subjects. Out of research on teaching and learning, using the new AI technology, we may see a revolution in the educational process quite comparable to the revolution in medical services that followed the deepening of our knowledge of the biological bases of disease.

## CONCLUSION

I should like to conclude by sharing with you my own vision of the future of artificial intelligence and cognitive science. I see no limits to the range of human thinking capabilities that can be embraced by these fields. I believe that, in time, the boundaries of AI will encompass all of human cognition, and I have tried to present here some of the evidence that leads me to this conclusion.

The boundaries are going to be enlarged gradually, not suddenly. The pace of application will be even more gradual, for it will call for enormous amounts of capital investment in hardware and especially software. Expert systems, and other schemes for thinking and planning, will probably move faster than robotics. Building systems that can sense and act flexibly in complex natural environments is still the harder task.

We should not suppose that the spread of automation implies that human resources must be unemployed. Our society has many needs that we do not feel are being met adequately. We are now engaged in a great national debate to determine how scarce resources are to be allocated among national security, health, consumer goods, investment, the environment, energy needs—all of these in the face of a growing national debt. If we have unemployment, it surely is not because we are too productive. And the correct measures to combat unemployment are surely not measures that will check the advance of our productivity.

I do not mean that a society undergoing change does not create hardship for many of its members. We must do better than we have in the past to permit our society as a whole to absorb the transient costs of beneficial social change. But we must do this without making change and progress themselves impossible or difficult.

I have proposed that there are two potential kinds of payoff from artificial intelligence and cognitive science. In my personal judgment, the biggest of these lies in gaining a deeper knowledge of ourselves and in improving our own processes of thinking and learning. There are many problems in the world today that we often think of as technological in nature: overpopulation, the bomb, stress on the environment, scarcity of resources. At the most fundamental level, of course, these problems are not technological. They are the problems of ourselves. They will only be solved as we learn to think better, to plan better, to cooperate better. AI and cognitive science, pursued vigorously, can make an important contribution to improving the use we make of our own minds in addressing the world's critical problems.



## CURRICULUM VITAE

Nobel Prize recipient Herbert A. Simon was born in Milwaukee, Wisconsin, and educated at the University of Chicago, where he earned his A.B. degree, Phi Beta Kappa, in 1936 and his Ph.D. in 1943. He began his teaching career at the Illinois Institute of Technology and had become Chairman of the Department of Political and Social Science there at the age of 33 when he moved to the then Carnegie Institute of Technology. He is now Richard King Mellon University Professor of Computer Science and Psychology at Carnegie-Mellon University.

Dr. Simon has been awarded honorary D.Sc., LL.D., Fil.D., and Dr. Economic Science degrees from many American and European institutions of higher learning, among them Yale University, Columbia University, the University of Michigan, the University of Pittsburgh, McGill University, Université Paul-Valéry, Lund University, and Erasmus University. He is a member of the National Academy of Sciences and served there as the Chairman of the Division of Behavioral Sciences of the National Research Council from 1968 to 1970. He is also a member of numerous other prestigious academies, societies, and associations in this country and abroad, including the American Association for Artificial Intelligence, the American Philosophical Society, the Society of Experimental Psychologists, the Psychonomic Society, Sigma Xi, the Organizational Science Society of Japan, the Société Royale des Lettres de Lund and the Yugoslavia Academy of Sciences. He is a Fellow of the American Academy of Arts and Sciences, the AAAS, the AEA, the APA, the ASA, the Econometric Society, and the International Academy of Management. He has, in addition, been visiting professor, lecturer, and consultant at colleges and universities and industrial and governmental organizations too numerous to list here.

Among his many publications are *Reason in Human Affairs*, *Models of Bounded Rationality*, *The Sciences of the Artificial*, *Models of Thought*, *Models of Discovery*, *Administrative Behavior*, and *The New Science of Management Decision*.

Dr. Simon received the Alfred Nobel Memorial Prize in Economics in 1978.