

Dr. Herbert Simon

Mr. President, today we honor Herbert Alexander Simon, the Richard King Mellon University professor of Computer Science and Psychology at Carnegie-Mellon University and the 1978 recipient of the Nobel Prize in Economics.

Professor Simon is a polymath who has distinguished himself in many fields. While earning his A.B. and Ph.D. degrees in political science from the University of Chicago, he began the study of decision-making in organizations that led to his seminal work, "Administrative Behavior." During his career, he has held academic and administrative positions at the University of California at Berkeley, Illinois Institute of Technology, and Carnegie-Mellon University, in departments of political science, industrial management, psychology, and computer science. He is the author or co-author of over 500 articles and over 20 books covering the wide spectrum of his interests. A member of the National Academy of Sciences and a fellow of many professional organizations, he has received many honors, including the Distinguished Scientific Contribution Award of the American Psychological Association, the A. M. Turing Award of the Association for Computing Machinery, and election as distinguished fellow of the American Economic Association.

In awarding him the Nobel Prize in Economics, the Swedish Academy of Sciences cited 'his pioneering research into the decision-making process within economic organizations.' His observations and research challenge the classical economic theories that business firms behave as mathematical optimizers. By focusing on how decisions are actually made in the economic arena, he has adduced that 'bounded rationality' leads to decision-making by 'satisficing' rather than by optimizing, and has

thereby opened a whole new approach to the study of economics.

Extending the scope of his inquiries from decision-making processes to problem-solving procedures, Professor Simon, in collaboration with Allen Newell and J. C. Shaw, pioneered the development of artificial intelligence with such computer programs as the Logic ^{Theorist} ~~Proving Machine~~ and the General Problem Solver. His subsequent research has employed information processing models of intelligence to explain human cognitive processes. In works for the layman, such as The Sciences of the Artificial and Reason in Human Affairs, he stunningly articulates a paradigm with which to comprehend the complexity of the behavioral sciences, thus drawing together many threads of his various lines of research.

President Kendall, I have the great privilege to present to you, upon the recommendation of the faculty of Gustavus Adolphus College, Herbert A. Simon, for the degree of Doctor of Science, "honoris causa."

The title of my particular address in this series of six lectures is "Some Computer Models of Human Learning." I picked this particular title because I think that in many ways concrete examples are the best teachers of what cognitive science and that strange area of inquiry called artificial intelligence. ~~what these areas of inquiry~~ are all about. But before I get on to my examples I think I need to say a few words in introduction and place my own remarks in the context of those that preceded mine yesterday morning. ~~I won't ask for a show of hands as to heard the talks yesterday morning, but I assume that many of you did.~~

One of the things that was very evident in the talks yesterday is that an area as complex as cognitive science can be investigated at many levels. In the addresses of Professors Edelman and Milner, we saw the investigation beginning at the level of the biological brain--~~beginning~~ at the level of tissues and cells. In the paper of Roger Schank we saw the investigation beginning from the top down--beginning with the analysis of complex human behaviors and in particular verbal behaviors using natural language.

I don't think we should be surprised that in an area as complex as this ~~one~~ there is more than one direction of approach or even that there should be more than one level of theory. The idea that we can take ~~the~~ complex behaviors ^{like} ~~we think of as~~ thinking, problem solving, use of language, ^{or} understanding, ~~that we can take all of these behaviors~~ and in one step of reduction explain them in terms of chemical and electrical events in neurons seems rather extravagant, rather ambitious. Instead, this field is following the example of many other natural science fields ^{by} ~~of~~ constructing ^{successive} levels of explanation where the task of scientists is to reduce phenomena at one level of complexity to explanation in terms of

concepts introduced at the next level down. ^H ~~And~~ ^H here we are, as I say, ^E
 in good company with the other sciences. Nobody supposes that biology
 can't advance until we have a complete understanding of quarks. Nobody
 even supposes that solid state physics can't advance until such time as
 we have a full understanding of quarks. And that, of course, is very
 fortunate. It is very fortunate that science can ~~be~~ sometimes ^{be} hung from
 skyhooks, because the foundations underground are often ~~very~~ mysterious
 and shifting, ~~indeed~~. And science often builds from both directions,
 from the top down and from the bottom up. ~~And so we have in my map of~~
~~this territory, and~~ I think each of the six of us has a slightly
 different map ^{of the territory, and} you can take all of these maps home and have an atlas, ^B but
 in my map of the territory there are very definitely at least two levels—
~~and perhaps three.~~ the neuron level at which very exciting work ~~that~~
~~you heard something about yesterday~~ is going on, and the information
 processing level, *which I hope to persuade you is equally exciting*

Just one short word on terminology. By information processing
 here, I do not mean exactly what Professor Edelman mean^S by it, ~~yesterday~~.
 He was using it as a label for ~~a particular point of view~~, a particular
 approach within a somewhat larger category than I will be concerned with;
~~so~~ when I talk about the neuron level and the information processing
 level, I am not making the same contrast that he ^{does} ~~was~~ between what he
 call^Sed information processing and the population approach to the field.

In the ~~developments of the~~ last twenty-five years, ~~since it~~ ^{has} been
 discovered that computers can do something besides number-crunching,
~~since it's been discovered~~ that they are quite general symbol-processing
 devices. ^{This} ~~that~~ discovery has been exploited in at least two ways. One is
 in the direction called artificial intelligence, which can be very

roughly defined as finding as many ways as you can to get computers to do clever things for practical reasons or just because it's fun, or a little of both. ^{fun?} But another direction which is more often nowadays labeled cognitive science, and with which I think we're more concerned here, ~~and that~~ is the direction of using the computer to understand the nature of intelligence, the nature of mind, including the nature of the human mind. ^{This is} ~~What it is, well,~~ really, the old problem of mind and body, -- what it is that enables a biological organ like the human brain to think.

~~Yesterday~~ Another distinction ^{is the distinction} ~~was made~~ that will be rather important in my discussion, between systems, ~~intelligent systems~~, whether they be computers or ~~whether they be human beings~~, ~~intelligent systems~~ ^{that} ~~which~~ perform intelligent tasks, (they play chess or they make medical diagnoses or they solve problems) ~~Those kinds of systems~~ and systems ~~which~~ ^{that} learn to perform tasks. Because there's certainly one important difference we know between computers and human beings: In order for a human being to be able to do anything, except ^{the} ~~that~~ limited range of things ^{that} ~~which~~ is evidently built in, ~~but in order for human beings to do anything~~, we have to learn. And since nobody knows our programming language, the internal programming language, the only way we can learn is by being exposed to experiences and to symbols from outside and, somehow or other, ^{going} ~~or~~ through a process of transforming those stimuli ~~and~~ ~~symbols~~ in such a way that the internal state of the system is changed so that we're able to do new things. ^{is} The process we call learning, a very impressive process if you look at its cumulative effects over a lifetime or ^{even} ^{although it is} over a childhood, sometimes a frustratingly slow process if we realize that a large number of us are still in school after having started ~~this process, with~~ the formal process, at age six, and the ^{for}. More exactly, ^{performance} ~~behavior~~ of a computer may be called artificially intelligent if the corresponding ^{performance} ~~behavior~~ of a human being would be called intelligent.

informal learning process from the time we drew our first breath.

Something must be wrong with a process that has taken us all these twenty-odd years to get us where we are. ^{and} So I think we have ~~this dual~~ ^{a ambivalent}

feeling about learning, that on the one hand it's very impressive that ~~this~~ ^{it} can be done ^{at all}, and on the other hand, that it is sometimes ~~painstakingly~~ ^{painfully}

slow. With a computer, ~~on the other hand, of course~~ ^{however} you can open the lid, ~~you're allowed to open the lid~~ of the box and put things in if you

know what to put in. And so computers, on some dimensions at least, can

become instantaneously intelligent. They can become instantaneously

pretty good chess players. Human beings probably could not reach the

levels of the best programs today in less than five years of devotion to

the game of chess, and most of us might never reach those levels, even

though we play weekend chess all ~~of~~ our lives. ~~I will~~ ^{the} focus of my

remarks will be on learning, but if we look at the strategy of research

in cognitive science over the past twenty-five years, we ~~will~~ ^{will} see that the

strategy has been ~~first~~ ^{initially} to try to understand human performance. ^{First,}

to try to understand how it is that a human being is able to solve a

problem. Having understood something about that, then ~~people~~ ^{researchers} are

beginning to go back and ask, "Can we conceive of mechanisms, ~~can we~~

~~conceive~~ of processes, ^{that} which could be implemented in something like a

brain and ^{that} can be implemented in a computer? Can we conceive of ~~such~~

processes ^{that} which could arrive at this state of performance, ~~which~~ ^{-- that} could,

in fact, learn and not simply be programmed?" ~~And~~ I will try to give

examples of ^{such processes} ~~that~~.

The Physical Symbol System Hypothesis

Now what I will say proceeds from a hypothesis. It's rather

pretentiously called the Physical Symbol System Hypothesis. Unless you

have a ~~rather~~ formidable name for a hypothesis, nobody will pay any

See
heading

attention to it, so we ^{will continue to} call this one the Physical Symbol System Hypothesis. By a physical symbol system I simply mean some kind of a device ^{that} ~~which~~ can deal with patterns, ^{where} ~~and~~ a pattern is ~~almost~~ any arrangement of things. There's the pattern of a human face, which allows us to recognize it. There's a rather symmetrical pattern of the trestles that hold up the ceiling of this room. So by pattern I simply mean some kind of an arrangement ^{that} ~~which~~ we can ~~presumably~~ learn to discriminate from other and different arrangements, or that we can learn to compare with similar arrangements, just as we can see the regularity of these trestles. [¶] What can we do with patterns? ^{For one thing, both} ~~We now, being human~~ beings, and also it turns out, computers, (patterns) can ~~be~~ read, ^{that} is, ^{patterns} ~~they~~ can be ingested from some external sensory resource. They can be ^{ital} written, ^{ital} that is, patterns can be created in the environment ^{by} ~~from~~ the system itself. They can be ^{ital} stored in what we call memory; ^{ital} they can be kept around for a long time, more or less accurately and presumably with more or less reliability. Patterns can be put in relation to each other ⁱⁿ ~~in~~ classical psychological terminology, they can be associated, and similarly, in ^{non-human} ~~other~~ pattern devices, we can build up relations so that one pattern points to another pattern in the system. As I've already suggested, patterns can be compared; I can compare ^{my two hands,} ~~those two patterns,~~ ^{by putting together my fingertips,} match them ~~together~~ and see that they are instances of the same pattern, -- the human hand, more or less similar instances in this case. You can have ~~strict~~ strict tests of comparison or looser tests of comparison. I guess ^{the} ~~that~~ test ^{I just applied to my hands} didn't notice that one was a left hand and one was a right hand.

What is the Physical Symbol System Hypothesis? The hypothesis is ^{this:} ~~that~~ The necessary and sufficient condition that a system be capable of

behaving intelligently is that it be a physical symbol system. The necessary and sufficient condition that a system be able to solve problems to think, to use language, to understand, ~~the necessary and sufficient conditions~~, is that it be able to read symbols, (that is, sense them), to write symbols, to set up structures of related symbols, to store those structures in memory, to compare symbols to see whether they're the same or different, and to behave differentially on the basis of whether those symbols turned out to be the same, or ~~turned out to be~~ different.

There are several corollaries to the Symbol System Hypothesis. The first corollary is that if the hypothesis is true, then it follows that computers can be made to behave intelligently, ~~because if the Symbol System Hypothesis is true~~, we know that computers, we can find that out by opening the box, so to speak, we know that computers are physical symbol systems. They have all of the simple capabilities that I mentioned of reading and writing, ^{and so on --} I won't go through the list again. ~~You know the list now.~~ So if the hypothesis is true, then computers can be made to be intelligent. [¶] The second corollary is that if the hypothesis is true, if this is the necessary condition for intelligence, then human beings must be at least physical symbol systems, because human beings are capable of behaving intelligently. [¶] I haven't defined intelligence, by the way. I simply am using it in ~~a common sense~~, in its common sense meaning. We all know how we judge our friends to be intelligent, ~~and of course all of our friends are intelligent, sometimes more intelligent than others~~, but we ^{also} have all sorts of ^{formal} tests, ^{which} ~~and these tests~~ generally consist of giving someone a task and seeing how successfully that task is ^{handled} ~~confronted~~. Intelligence is no unitary thing, because people can

be very intelligent along some directions, ^{and} ~~can be~~ very unintelligent
 along other dimensions. ^{is} So that ~~is~~ all I mean by intelligence, and when
 I talk about a computer being intelligent, I'm going to apply to it the same
 kinds of tests I will apply to the human being. I will judge its
 intelligence by the range and kinds of tasks that it can perform which,
 if performed by a human, we would say required intelligence. ~~Let's say~~
~~perfect fairness in the tests that we're going to apply.~~ ^{It} So that's the
 hypothesis--that human beings must be physical symbol systems and that
 computers can be programmed to behave intelligently. No one needs to
 accept the hypothesis or reject it in any hurry; there's plenty of time
 to consider it. It's an empirical hypothesis. We test it by looking at
 empirical facts, ^{running} We test it by ~~designing~~ experiments. For example, if
 we want to know whether a computer is intelligent in being able to per-
 form a certain kind of task, we can try to write a ~~computer~~ program that
 will enable the computer to perform tasks of that class, and as wide a
 class as possible. ^{Test the hypothesis that} If we want to ~~know whether~~ people, in fact, are
 intelligent because they behave as physical symbol systems--because they
 have these properties of manipulating symbols in their brains, ^{these}
~~capabilities; if we wanted to test that hypothesis, it's a little bit~~
 more complicated. ~~And~~ ^{The} usual strategy of testing that's been
 adopted ^{is} in cognitive science in the past twenty-five years ^{is}
 something like this. One writes a computer program, ^{since} ~~we~~ already know
 that the computer is a physical symbol system. One tries to write such
 a program so that it can perform some intelligence ^t task that we might
 ask a human to perform, and ~~it~~ perform it in a demonstrably human way.
 Not fly as an airplane flies, but fly as a bird flies ^t (if we want to
 understand birds). Not play chess by trying millions of combinations,

H. G. M. L. P. T.
H. G. M. L. P. T.
H. G. M. L. P. T.
H. G. M. L. P. T.
H. G. M. L. P. T.

In sum, we try to describe symbol manipulating processes (the program) that can simulate closely, and account for, human intelligent performance of the same tasks.

which a computer can do (and the best programs do), but play chess by looking very selectively at the important features of the board, which we know to be the way a skillful human being plays chess. ~~We can try to~~

~~write computer programs and then do with them what we do with any~~

~~A~~ ^{such} a computer program is a theory of the human behavior, theory. How do we test ^{the} theory? We let the theory generate its

consequences. In this case, we let the theory, the computer program,

generate a stream of behavior. We let it ~~play a game of chess~~ ^{make} or a move

in chess. ^{or} We let it solve an algebra problem. And we see whether, step-

by-step, it seems to follow the same kinds of procedures and make the

same kinds of errors, and get into the same kinds of backwaters that a

human being does in solving ^{the same} difficult problems. ~~And that~~ ^{This} kind of

evidence has been accumulating for about twenty-five years now about a

limited range of human tasks. We still are far from testing ~~this~~ ^{the}

hypothesis over the whole range of things ^{that} human beings do and call

intelligent. Professor Schank, yesterday, called to our attention

some of the difficulties in simulating, by computer, human behavior in

areas like everyday use of language. There are indeed such difficulties,

and there's plenty of research to be done in continuing to test this

hypothesis. There will be work for all of you who are motivated to join

this game. There will be work ~~for you~~ to do for several generations,

I'm sure.

^{have said,} As I ~~say~~, the early activity in this field ^{was} ~~has~~ mostly ~~been~~ directed

to performance. ~~To~~ getting computers to play chess, ~~to~~ make medical

diagnoses, ~~to~~ solve ^{Tower of} the Hanoi problem, or getting the missionaries and

cannibals across the river, whatever your favorite puzzle happens to be.

But in recent years progress ^{toward understanding intelligent performance} has reached a point, and enough ideas have

~~been generated about possibilities for the learning process that there's~~ ^{has encouraged}

^{in cognitive science about}
~~been a very strong renewal of research in the area of learning, in the~~
~~area of cognitive science.~~ That research has ^{produced} ~~resulted in the develop-~~
~~ment of~~ a number of computer systems that are capable of learning, some
of them in a quite humanoid manner, others in manners that may or may
not bear any close relation to human learning, ~~and I'd like to mention~~ In
the remainder of my talk, ~~here~~, I'd like to refer to three of those
^{systems,}
~~examples.~~ I picked these three as examples that I'm especially familiar
with. They are not in any sense a random or even a representative sample
of work in this field.

Subheading
Role Verbal Learning

~~First of all, psychologists~~ ^{psychologists} For the first fifty years or so of the
history of ~~that field,~~ ^{psychology} from the time of Ebbinghaus and Wundt in Germany,
psychologists were ~~very~~ fascinated with the way in which people did
verbal learning. I guess In those days, ~~in school, well, I'm sure in~~
~~those days in school,~~ children still had to learn lots of poetry, ^{and} they
~~obviously had to learn~~ lots of foreign language vocabulary. I am told
that children don't memorize nearly as many things in school ^{now} ~~anymore~~ as
they did in days gone by, and I don't know quite what to think of that.
I guess it's a good thing. ^{At} At any rate, psychologists became preoccupied
with these kinds of verbal learning processes, ^{note} ~~Even wrote~~ learning pro-
cesses that ~~is fact,~~ constituted a large part of school activity. And
there's a vast body of psychological literature which describes what
human beings do when they're faced with a verbal learning ^{task!} ~~situation,~~
~~with a problem.~~ ^b Of course Ebbinghaus wanted to do this really scien-
tifically; he wanted to control variables, so he invented something
called nonsense syllables that didn't have any meaning, ^{he} ~~and~~ did most of
his experiments with nonsense syllables. But we won't get into
nonsense, that kind at least, this morning. We'll talk about learning
in general.

There exists, and has in fact ~~has~~ existed for ~~now over~~ ^{five} twenty years, a computer program ~~It's generally~~ known as EPAM, which may ~~or may not~~ have been intended to stand ^s for Elementary Perceiver and Memorizer. ~~Some people think it stands for Epaminondas, but I don't know. You can~~ take your pick. There's a program called EPAM ^{EPAM} which, in fact, is capable of learning ^{learning} either a series of ^{items} ~~things~~, the way we learned the alphabet, or learning ^{associations between pairs of items,} ~~paired associates of things~~, the way we might learn a foreign language vocabulary. There's no claim that EPAM understands semantically the things that it learns. ~~What EPAM is~~ is an attempt to model ~~certain aspects of~~ the human learning process when humans are learning things more or less by ~~what we call~~ rote. ^{It} Today we can explain ^{most} ~~a very large part~~ of the main phenomena that were discovered in the verbal learning laboratory ^{ies} of psychologists. We can explain those phenomena with the EPAM program in the sense that if you give EPAM the same materials that were given to the human learners, ^{it} ~~they~~ will produce the same kinds of phenomena. For example, it has been many times demonstrated that syllables consisting of three unrelated letters, real nonsense syllables, take about three times as long to learn as simple one ^{meaningful} syllable ~~English~~ words, or whatever your native language is, words. (It could be Swedish words, I guess, ^{here} ~~here~~, or English words, or Dutch words, or take your pick.) But three times as long. ^{Without} ~~Why that~~ ^{trying to explain here how EPAM does it,} ~~mysterious number three? I don't want to go into the anatomy of EPAM,~~ ~~but~~ here's a program that predicts, in fact, that English words will be learned about three times faster than three-letter nonsense syllables. Various phenomena with regard to how the ^{a dissimilarity} similarity between items that are being learned ^{also} affects the learning ⁱⁿ in some cases it facilitates it and in some cases it interferes with it. EPAM can predict the major effects along those dimensions, ^{also}. The mechanisms within EPAM that cause it to model the human behavior so closely constitute a plausible theory of the processes of human verbal learning.

Short-term Memory, Recognition and Attention

Out of the research on the EPAM model and similar models, combined with laboratory research ^{using} ~~in which we study~~ human subjects, ~~the~~ two important ideas began to emerge. ~~Now~~ ^{One} of them was around before EPAM ~~was~~ ^{then}. ^{This idea} In fact, introduced by George Miller in 1956. ~~One~~ ^{One} was that it is terribly important in human performance ~~is the fact~~ that we have ~~something called~~ a focus of attention, otherwise known as a short-term memory, ^{that} everything we pay attention to, ~~that~~ everything we do that requires some modicum of attention, ^{when} has to be squeezed through the bottleneck of our short-term memory. ^{We} add a column of figures, those figures, one by one as well as the running total, have to be in short-term memory. ^{Short-term memory} ~~And this~~ ^{is an extremely} narrow bottleneck, and ^{it} turns out to be one of the most important features of the architecture of the human thinking system in determining the rate at which we can think and ~~in~~ ~~determining~~ the way in which information has to be organized ^{so} ~~in order~~ that we can carry out complex sequences of thoughts. Today, we speak of the short-term memory, ~~or some of us do, as though it were, sometimes we~~ ^{no more than six or seven} speak of it, as though it were a little box which had ~~six~~ little compartments in it. ~~Some people think it has seven compartments in it. We won't~~ ^{in which} quarrel about one compartment more or less, and that we can hold ~~in that~~ ~~short-term memory~~ something like a half-dozen familiar items, where a familiar item might be anything ~~which~~ ^{acquainted} you have become ~~familiarized~~ with. For people in this room, familiar items might be any one of the fifty thousand English words that are in your vocabulary. ^{(Perhaps you} ~~You maybe~~ ^{don't} know that you have fifty thousand words in your vocabulary, but this ^{quantity} ~~this~~ has been estimated ~~for you~~, and it is of that order of magnitude. It may actually be twice that number, but we won't quibble about factors of two here.)

~~You can hold any six of these in memory long enough.~~ If I gave you a list of six ^{words}, you can hold them in memory long enough to repeat them back to me. But watch out if I give you nine. You're going to be in trouble. ^{There is} A vast body of experimental evidence which is ~~in turn~~ compatible with the ~~SRM~~ ^{of a limited short term memory} model and the effects of which are predictable by the ~~SRM~~ model.

We think today that the structure of the short-term memory has a very great deal to do with the way in which knowledge is stored in the mind of an expert. We believe that a great deal of what the expert is able to do, he or she is able to do because of an ability to recognize cues in a familiar situation, ^{recognize cues in any situation within} the domain of expertness, and having recognized those cues, to use ~~those~~ ^{them} ~~eyes~~ to access relevant information in long-term memory.

It doesn't surprise any of you ~~if you walk down the street~~ that when you see the face of a familiar friend, you immediately recognize that face. Moreover, ^{though} ~~you can probably~~, not all of us are real good at this, ~~but~~ you can probably retrieve from long-term memory the name of the friend, ^{or} If not the name, a lot of facts about the friend. ~~We think~~ ^{This} ~~that~~ process of recognition followed by retrieval from long-term memory underlies most human expertise. We even have estimates today of how many ~~of these~~ ^{chunks} the same familiar little chunks you store in short-term memory, ~~how many of these chunks~~ it requires to make an expert. The answer again comes out, ⁱⁿ order of magnitude, ^{to be around} fifty thousand. ^{For example,} A chess grandmaster can recognize fifty thousand ^{or more} different kinds of features and patterns of pieces that are likely to occur on a chess board during a game.

~~I'll just mention,~~ ^{provide} Just to ~~give~~ ^{with} you a sample of the kinds of

experimental evidence that buttress up hypotheses like this, I'd like to give ~~you just~~ a short description of a classical chess experiment. You take a chess board, eight-by-eight squares, and arrange on it the pieces from a well-played game, ~~from oh~~ ^{at} maybe the twentieth move, ~~the game~~ ^a that's well-played but unfamiliar to the subjects ~~you're going to use~~ in the experiment. You ~~put~~ ^{display} this position, ~~you display it~~ to your subjects for five to ten seconds, and then withdraw it; and ^{than you} ask them to reproduce the chessboard. If ~~your~~ ^a subject is a master or a grandmaster, the ~~board~~ ^{pieces} will be replaced almost perfectly, better than 90% accuracy, even though there are 23 pieces, on average, on such a board. If the player is something less than a master or an expert player at least, he or she will be lucky to get six pieces back on the board ^{correctly} ~~right~~. Oh, you will say, ~~You need~~ special kinds of eyes, special kinds of visual imagery, to be a chess master or grandmaster.

Well, let's take the next step in the experiment. ~~The next step, we~~ ^{at} simply take the same pieces and arrange them at random on the board and expose them for five ^{to ten} seconds. And now the ordinary chess player replaces his six pieces on the board just as he did before, and the master or the grandmaster replaces about six pieces on the board. ^{It has} ~~Nothing about~~ the eyes, ^{but} a great deal to do with how many hours have been spent in staring at chess boards, ^{misspent} ~~misspent~~ youth you might say, ~~Well, it depends on how you feel about chess.~~ How many hours have been spent in staring at chess boards? How many patterns have become familiarized and have been stored away in long-term memory, associated with information ~~also~~ about what to do when you see ~~such~~ ^{is seen on the board} a pattern? And so we ~~see~~ ^{observe} the expert in every field of expertise not only recognizing familiar friends in the situation ^{but also recalling all sorts of information about these friends} he sees, the doctor looking at your spots ~~or whatever it is~~.

you come in with and say, ^{ing} Oh yes, ^{that's measles} ~~but~~ forming an immediate hypothesis, which may be subject to revision, ~~an immediate hypothesis~~ about what might be going on in the situation. And so we think today that ^{this} ~~that~~ kind of performance ^{— this} ~~that~~ kind of expert intuition if you want to call it that, or insight, [—] can be given a perfectly reasonable explanation in information processing terms, in terms of ~~these~~ recognition processes.

Solving Algebraic Equations

Let me go on to another kind of learning, because all of us, particularly those of us ^{here} who are teachers ^{here}, and I hope most of the students, know that rote learning except for ~~very~~ special purposes really isn't a very good thing. We don't encourage it in our students. Rote learning is great to get through the Friday quiz, but by Monday it seems all to have evaporated again. We really want to talk about learning with understanding. What does it mean to understand what we've learned? Roger Schank pointed out to us yesterday ^{that} the word "understanding" may mean many different things, and I'm only going to touch ~~base~~ on one or two of them.

All of us here, at some time in our life, ^{has} had occasion with greater or lesser pleasure to learn to do algebra, at least simple algebra. We learned to solve equations. The teacher would write on the board, $3x + 4 = x + 10$. ~~$3x + 4 = x + 10$~~ . If you have paper and pencil, you can solve that; I'm not

going to ask you to. We won't grade you at the lecture here. ~~$3x + 4 =$~~

~~$x + 10$. Some of you here probably still remember what to do about that,~~

Probably most of you here ^{still} remember what to do ^{about such an equation} ~~about that~~. In a moment

or two, you ^{will} ~~would~~ have solved ^{it} ~~the problem~~. Probably fewer of you could

^{But suppose} quickly, ~~if~~ your child came up and said, "Daddy or Mommy, I'm having

trouble with my algebra tonight. Will you help me?" Probably fewer of you ^{would} ~~will~~ have immediately on the tip ^{clear explanation} ~~of your tongue a set of ideas as~~

Separate line →

~~of~~ exactly how the trick is done. Maybe you can do it yourself, but do you know what ^{you are doing} it is while you are doing ^{it?} ~~the trick?~~ There's been a good deal of study now of how people solve simply algebraic equations--a simple skill, but like ~~an awful~~ ^a lot of ^{the} skills that we use in our school courses, particularly but not exclusively in the areas of science and mathematics. ^{the process} In its broadest terms, ~~it~~ goes something like this. (I won't even refer to algebra in the first version.) I am in a certain situation, standing in front of this podium. I need to be in a different situation. Eventually I'm going to have to walk down and get out of here. Now, I have to ask, what's the difference between where I am and where I want to be? Well, in this case it's a difference in position or in place. What do I know? What knowledge do I have which tells me how to reduce such differences? Well, there are bicycles, there are taxis, there's something called shanksmare. Bicycles and taxis aren't very appropriate here, it would be a little hard getting them down the stairway, it might be hard getting them into the hall; but I always do have my legs. And so we've solved this very simple problem. When the time comes, I'm going to walk down--a very deep solution. It takes ^{babies} ~~kids~~ ~~a while to, a couple of years before they're~~ ^a year or so before they're able to do that.

The thing I've been describing with this ~~simple~~ stupidly simple, example is something called means-ends analysis. I'm here, my goal is ^{to be} there; I detect a difference between where I am and the goal; ~~I reach into memory;~~ I use that as a cue to reach into memory to see whether I have any experience in reducing ^{differences} ~~experiences~~ of that kind. Yes, there is such a means of reducing differences, and I apply it. I am now closer to my goal, maybe I'm at the head of the stairs. Now I have to have a new difference-reducing means called walking down stairs, etc.

How do we do algebra by means-ends analysis? Very simply. I have an expression like $3X + 4 = X + 10$; what do I want? Well, I'll have a solution if I have an X ^{followed by} an equals sign, ^{followed by} $X = N$, and a number. ^{Of course,} it has to be the right number, ~~and I won't go into that~~. It doesn't pay off if you just write down any number; ~~that would make the problem too easy.~~ I need an X , an equals sign, and a number.

How do I change an expression like the one I have into one that consists of just ~~an X , an equals, and a number?~~ ^{$X = N$?} Well, we do have all kinds of operators. We're allowed, without ~~changing~~ ^{altering} the value of ~~that~~ ^{X} number, ~~we're allowed~~ ^{we may} to subtract ~~a number~~ ^{a term} as long as we do it from both sides. ^{of the equation,} We're allowed to add ~~a number~~ ^{a term} as long as we do it to both sides, ^{We may} multiply both sides by the same ~~number~~ ^{term and}, divide both sides by the same ~~number~~ ^{term} (except, watch out for zeros). That's about all we're allowed to do, isn't it? Those are the only actions. ~~I can walk, I can climb stairs, that's about all I can do.~~ ^{We} I can add, subtract, multiply, and divide the same quantity from both sides of the equation. ^{If} If I have an equation which has not only X 's but numbers on the left side, ^{and} not only numbers but X 's on the right side, and I want to reduce it to an equation with just an X , ~~and an equals~~ ⁼ and a number, let's get rid of some of those differences. Let's subtract away the numbers on the left side, which I don't want in the final result; ^{then} let's subtract away the X on the right side, which I don't want in the final result. ~~Let's divide by the $2X$ which,~~ ^{a $2X$} If you've worked this out, you will find you have ^{on} the left side by now, ^{so} let's divide through by the two, and we will end up with an X and an equals and a three, assuming I did it right. An X , an equals, and a three. We can check whether I did it right, ^{by} and sticking ^{the answer} that back in the equation. ^{The whole process looks like this:}

$$\begin{aligned}
 3X + 4 &= X + 10 \\
 3X &= X + 6 \\
 2X &= 6 \\
 X &= 3 \\
 3 \cdot 3 + 4 &= 3 + 10 = 13
 \end{aligned}$$

Today ^{that proceeds in this way}
 We know how to describe a system ~~like that~~ in terms of computer programs ~~today~~. We know how to describe it perfectly well, and when we do describe ~~it in that way~~, it looks very much like the processes that enable a person to solve this equation, ^{the way they} looks very much like the processes that allow the physician to be an expert diagnostician, and that allow the chess player to be a grand master. ~~chess player~~. In the case of the physician, "See those spots, guess that it might be ~~XX~~", suggest the following tests. If you're the chess player, "See that open file, consider putting a rook on it. Well, ~~don't~~ don't do it in a hurry, think about it awhile, but consider putting a rook on it." If you're solving algebra problems!
 If ~~you have an~~ ^{the} equation ~~and it~~ has a number on the left-hand side, consider subtracting that number from both sides. ~~A system~~ If you've studied ~~the~~ stimulus-response psychology, behaviour ^{it} psychology, you can think of ~~this~~ ^{such a procedure} as a stimulus followed by a response. If you studied computer science instead, you can think of it as something which nowadays we call a production. ~~A~~ condition or cue, followed by an action! And ~~our~~ ^{If condition C is satisfied, then take action A: C → A.} hypothesis ~~here is~~ that most human skill is embodied in the brain in the form of productions, ~~a~~ perceptual capability of responding to a cue followed by retrieval of information about the action to take.

The ~~expert~~ ^{person} who does this algebra problem ~~and algebra problems like~~ doesn't have to have fifty-thousand chunks. ^{in order to become} (Probably ~~before becoming~~)
 a Ph.D. mathematician ~~that~~ ^a person has to acquire the fifty-thousand chunks, or ~~maybe~~ ^{even} a hundred-thousand chunks, ^{But} to do algebra problems of this degree of difficulty, ~~and mind you learning to do that is~~ ^{learned from} about one week's work in the algebra course, ~~the set of chunks that you need, the~~ ^{you need a} set of productions, ~~you need to do that is in this case just five in~~ ^{just five} number. I've already ~~said~~ ^{sketched for} them ~~to you, several times~~. Any system

separate level

I don't understand this at all!

~~can solve simple equations if it~~
 including a computer or a human brain ~~that~~ ^{and applying} has a way of storing ~~in it~~
 the following set of procedures--if there's a number on the left,
 subtract it from both sides; ^{if} if there's an "X" on the right, subtract
 that from both sides; ^{if} if there's a number multiplying "X" on the left
 and the number isn't one, divide both sides by ~~that~~ ^{it}. ^{And} if you have an
 expression that looks like " $X = N$ ", halt and check. ^{Four} For productions
~~will~~ ^{will} do a whole range of ~~those~~ ^{these} algebra problems.

~~If that does, in fact, have any resemblance. Now, let me make one~~
~~previous statement to that. I do not mean by that that anyone supposes~~
 that you can learn to do algebra by memorizing those four rules. When I
 say those ~~four~~ rules have to reside in the brain in order for you to do
 algebra, ~~what I mean by that is~~ ^{that the perceptual skill} of
 recognizing the cues ^(the "if" conditions) and ~~the access to memory that follows that percep-~~
~~tual skill, the access to~~ ^{accessory in memory} the knowledge of ~~the~~ ^{what} actions to take, ~~has to~~
~~be available in memory.~~ ^(the right hand rules of the productions) For most of us, at least those of us who have
 continued to use our algebraic skills, those four productions are so
 well stored ~~there~~ ^{they} that ~~no longer do they~~ require conscious attention or
~~very~~ much conscious attention for their application. I don't have time
 this morning to go into the question of where the boundary of conscious-
 ness ~~is~~ ^{but} ~~certainly there are processes that automate~~ ^{the application} ~~the application~~
 of a great many of these productions, ^{so that a simple recognition}
~~takes place of a situation~~ ^{occurs} and an action follows it. ~~That's that's~~
 characteristic of expert behavior. ^{Such "intuitive" responses are} ~~If that does~~ correctly characterized
 or even partially correctly characterized the way some skills are held
 inside the human skull, then it becomes an interesting question of how
 learning ^{of the skills} ~~might~~ takes place. How might a person acquire such production
 systems? Again the computer comes to our aid because today we know how

Put each
production
on a
separate
card

Subheading

Learning
algebra

to program so-called adaptive production systems. There's nothing very formidable about an adaptive production system. An adaptive production system is simply a computer program which consists of productions, and you can write as general a computer program as you like in this particular format it turns out and create a Turing machine out of productions. So, an adaptive production system is simply a set of productions which that includes a subset ~~that are~~ capable of constructing ~~or~~ generating new ^{symbol structures} ~~symbolic expressions~~ in the form of productions. After all, ~~these~~ productions must be stored in the brain in whatever symbolic language the brain is using, and what an adaptive production system does is to create new symbol structures ^{new productions,} of exactly the same kind ^{and} put them into the system. Once they are in the system they are operative, and can run along with all the other ~~things~~ that are there already.

How might that work with the algebra example? Suppose we had a clever student, and this idea was ^{initially} suggested ~~in the first place~~ by watching ^{other} clever students. ~~suppose we have a clever student~~ who encounters this chapter of the algebra textbook for the first time. One thing the student ^{might} ~~may~~ do is to read it through very thoroughly--very unlikely. Another thing a clever student might do is page through and see if there are some worked-out examples. Surely there will be some, ~~worked-out examples in that chapter.~~ Then the student will study ^a ~~the~~ worked-out example very carefully, study it in the sense of--well, what makes it work? What makes it tick? ~~Why does it tick?~~ After a little while that clever student may be able ~~to work simple algebra problems, may be able~~ to solve simple linear equations. We see that happening all the time with our clever students in school. They do a lot of their learning from worked out-examples, Sometimes worked-out examples in the classroom on the blackboard, very often worked-out examples in the textbook.

Well, what is going on there? You see, all the information that is needed to construct the production system ^{for solving such} ~~that we used to solve that~~ problem is present in the ^{textbook} ~~example, on the page~~. Because the student can compare successive lines of the solution, can see what changes have been made and by comparing those changes to the final result ^{" $X = N$ "} ~~that X equals a~~ ~~number~~, can see the motivation for ^{the} ~~these~~ changes, can see the gradual reduction of difference ^s between the original expression and the final expression. (See the ~~example~~ ^{illustrated above}.)

As you can see, I'm waving my hands, which in computer language means I'm ^{passing} ~~going~~ over all the difficult issues, but there are running programs today ^{that} ~~which~~ do this, ~~and help to keep us honest in describing this kind of research~~. Essentially by using ~~the same kinds of~~ means ^{ends} and analysis but using it in a questioning mode. ^{Such and such an action was taken, the person moved over there, and why?} Let's compare ^{the situation} ~~it~~ with the goal that was being aimed at. Let's now define a difference in terms of how we ^{were brought} ~~got~~ nearer to that goal by that action. Let's associate that difference with the action that was taken, ^{and} ~~Now~~ we have a new production. We stuff it into the system. It's ready to go the next time ^a ~~an~~ ^{similar situation} ~~algebraic equation~~ is encountered.

Can this work with human beings? ^{Can you} ~~Another game you can play, then, is~~ to construct teaching material on this basis; ~~You can construct some~~ ^{it has been done} ~~problems~~. In fact, ^{now is from an experiment where} ~~The example I'm going to give you is not this algebra~~ ~~one, it was the example of factoring quadratic expressions. The reason~~ ^{children were presented the task} ~~we didn't do the algebra but did the factoring and quadratic expressions instead,~~ ^{experiment with solving equations, but with} ~~it~~ had to do with ^{the} ~~a~~ particular point in the school year when we were ready to do the experiment. The students had already done ~~the algebraic~~ equations, and they were just coming on to factoring. We constructed a

set of written materials ~~only taking~~^{of only} three or four pages and consisting wholly of worked-out examples and problems for the student to work, ~~and~~^{carefully designed} ~~carefully designed~~^{The} worked-out examples, ~~so they would~~^{were carefully designed to} introduce the differences between the present situation and goal one by one. We presented these to classes of students. (We happened to have done this in China, but I think it ~~would~~^{will} work just as well with American students.) We'll find out in the coming year.) ~~We presented this to classes of students and~~ Within 20 minutes, 90 percent of the students in those classes were factoring quadratics. I don't know whether you think that's a great thing or whether you ~~did~~^{learned} that very easily when you studied algebra, but usually about 2 days of the algebra curriculum is devoted to ~~the task of~~ learning to factor quadratics.

I don't want to make too much of a single experiment that will certainly need to be repeated many times, ~~and~~^{and} ~~It~~ will need to be compared with other methods of teaching, ~~but~~ The point I wish to make is that not only can you build for a computer an adaptive production system that will teach itself to solve equations in algebra or to factor quadratics by looking at worked out examples, but you also can use this method for teaching human children, ~~and~~ It appears, at first flush at least, to be a not ~~completely~~ ineffective method.

The first comment we got from the teacher of the ~~first~~^{initial} class in which we ran such an experiment was, "and you didn't say a word to them." He evidently had the faith, ~~that~~^{which} all of us college professors ~~have~~^{hold,} that the way to spread the germs of knowledge is through the spoken word, by spraying the words on ~~at~~ our audiences. It seemed to him a little magical that the whole thing could be brought off without ~~one~~ words. Of course, there were written words, or at least written equations, ~~in the teaching materials.~~

The other point I wish to bring out by this example is that cognitive science has now reached a point in its development where it ^{is} no longer dealing just with toy tasks. It ^{is} no longer just dealing with getting those missionaries and cannibals across the river. It ^{is} dealing with school level tasks. For that reason, if enough excitement can be generated about this area of research, and I can't imagine why it ^{won't} generate excitement; ^{and} if enough bright people will come into this area, ~~research~~, I think it can be brought within a reasonable period of years into real relevance to the educational process, ~~and~~ I think that's an important task for us ahead.

Subheading

Discovery: Extracting Laws from Data

Let me turn to a quite different example. As we all know, computers can only do what you tell them to do. ~~See~~ Even these systems I've been describing were only doing what ^{we told} ~~you tell~~ them to do. An adaptive production system was simply learning from examples because ^{we} ~~you~~ told it to learn from examples. ~~Of course~~, If you think about it that way, ^{see that you} you can tell computers to do quite general things. How general? ~~can~~ you ~~tell them~~? How vague can the specification be? ~~The way to study~~ ^{this} ~~that~~ and ~~again~~ the way to get some hint as to how human beings ~~perhaps~~ attack vague problems [#] like many of the problems we deal with [#] in everyday life and many of the problems we deal with [#] in science [#] is to see whether we can, in fact, give a computer ~~program~~ more general tasks. I suppose we would think that the tasks of discovering laws ^{by seeing patterns} in raw data might be a task that would require a certain amount of independent thinking. Maybe we'd even want to use that sacred word "creativity" to apply to a human being who's able to take some data and out of ^{these} ~~this~~ data find order and ~~find hidden in the data~~ some kind of regularity or ~~other~~ scientific law. So, let's take that as our arena ^{of} study.

Let me give you an example first that comes not out of scientific laws but out of intelligence tests. ~~One standard kind of item, and~~

and you've all had it at one time or another. ~~One standard kind of item in~~ an intelligence test ^{is} the so-called Thurston ^E letter series completion task. You're given a sequence of letters and you're asked ^{what} comes next. Well, we'll all go through one together. A B M, C D M. What's next? ~~Oh, don't be bashful.~~ You all know the answer--E F M. Now there are some mathematicians here who will say, "anything could have been next." That's perfectly true, I could have started over again, A B M, or done anything; however, I would advise you that if you take such a test and want to get into college, you'd better answer E F M.

What was involved? ^{How did you} ~~What did you do in doing~~ that? What's this mysterious process of seeing a pattern? ^{It's} Not so mysterious. It's based on the elementary symbol processing procedures I mentioned at the very beginning of my talk. You all heard in the sequence I read the letter M--you heard it twice. ~~You heard the same,~~ ^{you} compared two symbols and they sounded the same, A B M, C D M. So you had an anchor point there. You could ^{periodically} conjecture at least ^{in the sequence} that perhaps ~~every so often~~ an M would appear.

Secondly, all of you know about the English alphabet. You know that B follows A, that D follows C, and therefore you know that a reasonable thing to follow the D is an E. ^{To solve the problem, you used your} ~~It's, in fact, the~~ knowledge ^{your} stored in memory of the alphabet plus ~~the~~ knowledge ~~stored~~ in memory about identity of letters, plus your ability to recognize situations where these simple patterns show up. ~~The fact that~~ ^{When one thing} follows another in orderly succession, the way the lights do on ~~the roof~~ ^{of} this ceiling ~~on the cards do across the table here,~~ that you are

prepared to notice that repetition, and you are prepared to notice that sequence in alphabets, in numbers, and in a few other series. In music, ~~you can notice the sequence of~~ ^{chromatic} ~~the diatonic and~~ ^{cycle} ~~grammatic~~ scales, the ~~circle~~ of fifths.

All of these elements of musical pattern that enable us to know whether we are listening to Bach, ^{or} Beethoven, ~~or whatever we are listening to.~~ Our ^{detection of pattern} ~~base~~ can be shown to be based on the ability to recognize repetition and the ability to recognize sequence.

So, here we have already discovered a very simple scientific law. We've learned that after A B M, C D M, there follows E F M. Let's consider laws that have had more historical interest than that particular law, unless that was the one you failed on the intelligence test, then it has historical interest for you I suppose.

Kepler looked at the heavens. It was very hard to look at the heavens in those days. They had very primitive instruments, but he and his predecessors were able to get some data about the distances of the planets from the sun and the period of time it took for each of the planets to revolve around the sun, ~~and similarly for the planets around Jupiter and around Saturn. At least they had that by Newton's time.~~ Kepler wondered whether there was a relation.

~~We have today, and we~~ found one, and it ^{is} called Kepler's Third Law. ^{BACON} We have today, a computer program called "Bacon," ~~it's~~ named, of course, after Sir Francis Bacon because it believes very deeply in Francis Bacon's ideas about induction. What the program Bacon does, following the advice of Sir Francis, ~~what the program Bacon does is to~~ look at data and try to induce from those data pattern. The first thing it looks ^{for} ~~at~~ in data ~~is to see whether there~~ are any correlations, ~~lying~~ around. In this case ^{of planetary motion} it would discover that as the distance of the

planet increases, the period of revolution also increases. What Bacon is looking for ~~as a goal out there~~ is, of course, an invariant--some kind of ^{law} variant law hidden in the ~~data~~ data.

So, If you have one ^{variable} ~~thing~~ that's increasing while another thing increases, one ~~kind of~~ primitive idea is ^{to} try dividing one by the other, ^{for} maybe the ratio will then be a constant, ~~and~~ Bacon does that--no luck, no constant. It gets a new quantity, P/D , ^{of revolution} ~~P over D~~ period over the diameter ^{of the orbit}. But it notices that ~~that quantity~~ P/D is also correlated with the diameter of the orbit, so it tries dividing again. It gets P/D^2 , ~~P over D squared~~--still no jackpot, still no invariant. It now notices that P/D^2 and ~~P over D~~ vary in opposite directions; as one goes up the other goes down. Well, maybe we'll have luck. Maybe if we multiply them together, we'll get an invariant. It multiplies them together, ~~it~~ ^{and} finds an invariant-- P^2/D^3 ~~P squared over D cubed~~, which is Kepler's third law.

Now notice that Bacon is trying different functions, but notice ^{also} that it's not doing it at all in a random way. And notice that, in fact, it finds the ~~correct~~ function after very little trial and error. I've ~~perhaps~~ oversimplified a little bit there, but essentially that is what it does, and does it in ^{less than} ~~something~~ under a minute on a ^{modern} ~~reasonable~~ computer.

no ~~It~~ ← Kepler took maybe ten years, but he was preoccupied with other things. His mother was being tried for witchcraft during part of that period and his life was rather complicated.

~~One other example.~~ I'm not going to ~~try to~~ give a whole long list of examples here of what Bacon does, ^{will serve} But one other example to illustrate ^{that} ~~that quite~~, let's call them natural processes, quite understandable processes, can be used by an information processing system to discover natural laws.

Suppose I had two objects here, and you have to imagine I'm a spring connecting these two objects. You have to stretch me a ~~little~~ bit,

leave line if you stretch the spring and then release ~~and~~ the objects will be accelerated.

~~Suppose~~ Suppose we can measure those accelerations. Bacon will rapidly find that the ratio of acceleration of Object A to ^{that of} Object B will be the same no matter how much I'm stretched or how little I'm stretched--the ~~same~~ ratio of acceleration. ^{will be the same,} ~~There~~ is a little scientific law. I don't know how exciting it is, but ~~there~~ ^{is} it is a ~~little~~ scientific law.

Take two different objects now, or take ~~this~~ Object ^A and a third object, and again we'll find the ratio ^S ^H ~~to~~ the accelerations are constant, but with a different constant. Now under ~~these~~ circumstances where Bacon finds a law that expresses a relation between two ~~things~~ ^{objects}, ~~under those circumstances, Bacon, I suppose this is a philosophical point of view it has,~~ Bacon will try to attribute a property to each of ~~these~~ object such that ~~these~~ ^{the} properties will explain the behavior of the ~~two~~ objects.

~~So,~~ ^{In} this case Bacon will assign, and I won't go through the procedure whereby it's done, but Bacon will assign a number ^{m_A} to ~~this~~ Object ^A, and a number ^{m_B} to ~~that~~ Object ^B, and then it will find that ~~this number~~ ^{m_A} multiplied by ~~its~~ ^{the} acceleration ^{of Object A} ~~times that number~~ ^{m_B} multiplied by ~~its~~ ^{the} acceleration ^{of Object B}.

~~Excuse me. This number, multiply this acceleration plus this number,~~

~~multiplied by the next acceleration~~ is always zero, which is an expression of the conservation of ~~mass~~ ^{momentum}. $m_A a_A + m_B a_B = 0$, ^{law of}

Notice Bacon did two things ^{that} here. First, it introduced new concepts. Nobody told it about ~~the~~ properties of ~~that~~ ^{the two} objects, ^{which object?} It introduced such a property ^{the} property which we know of as inertial mass. Secondly, by the use of ~~such new~~ ^{the} ~~process~~ of properties, it discovered ^a law. In this case, a very basic conservation law, of

physics, the conservation of ~~mass~~^{momentum}. Given other kinds of data, Bacon has succeeded in introducing the concept of specific heat in experiments involving mixing liquids at different temperatures, ~~different liquids at different temperatures~~^{that}. This was a problem which defied solution for about 30 years after the invention of the Fahrenheit thermometer.

Fahrenheit and ~~Boerhaave~~^{Boerhaave} Dutch chemists, got half way, but they ~~really~~^{were} stumped when it turned out that mercury absorbed less heat than water ~~for~~^{for} everybody knew that ~~since~~^{since} mercury was much heavier than water, it ought to absorb more heat, ~~than water~~. It was Joseph Black, the Scotch chemist, who introduced the idea of "specific heat." Bacon, knowing nothing about temperature or heat, ~~Bacon purely~~^{arrives, purely} by induction from the data, ~~arrives~~^{arrives} at the concept of specific heat, in order to find a conservation law for these phenomena. ~~It also finds~~^{introduces the concept of the} index of refraction. And it ~~can~~^{can} find many more, ~~well~~^{concepts and laws}. I don't want to go down the list of Bacon's wonders, ~~but~~^{for I} would like to make ~~just~~ one more point before I conclude this morning, but I think Bacon does demonstrate ~~is~~ that one can begin to ~~think about~~^{model and explain} those forms of human intelligence ~~which~~^{that} are usually supposed to involve genuine creativity. One can think about what processes it ~~could be that~~ would lead a scientist to discover a new scientific law. Now, inferring things from data is not the only way in which scientists infer new laws. I want to make that quite clear. But we can show historically that in the cases of Ohm, who invented Ohm's law of electricity, in the case of Kepler, in many other celebrated cases in the history of science, ~~in fact~~, there was no particular theoretical foundation for the discovery, ~~In fact~~, the discovery had to be an induction from fact because that's all there was. BACON shows ~~how~~ ^{how} ~~discovery~~ ^{discovery} of these kinds can be made by a physical symbol system.

TP 2

subheading
 Representation of Knowledge
 I need to mention one other area before I conclude here. That is

the area nowadays usually called representation. ~~Again~~ From human psychological experimentation we have good evidence that when someone takes a problem ^{presented} ~~given to him or her~~ in words and translates that problem into a set of equations and then solves the equation^s, or if someone is given a statement in the French language and translates that statement into ~~a statement into~~ the English language^s in both of those cases we have quite good evidence that the process is not a process of direct conversion from the input code or the input language into the output language, the mathematical ^{equations} language, or the English. ~~That~~ ^{There} intervenes between some ~~other~~ ^{input stimulus and output response} kind of internal representation, some other way of holding ~~that~~ ^{the meaningful content of the stimulus} information in our head, which we usually call ^{this intervening symbol structure} a semantic representation, of holding ~~the meaning of what is input in our head and that that~~ ^{what it} serves as a mediator between the input and the output.

You don't translate from French to English^s at least if you're any good at French and English, you don't. You translate from French to the meaning of the French ^(the semantic representation) and from the meaning of the French into English. It sounds like a slow way around, but it's really the only fast and correct way, ~~around~~ ^{supposed}. Similarly, ^{supposed} a physicist, or ~~again~~ a student in high school, ~~who~~ is given one of these problems about an 80% alcohol-and-water mixture which he wants to dilute into a 60% alcohol-and-water mixture. The students who are successful in solving ~~these~~ ^{such} problems do ~~it~~ ^{so} by means of an internal representation of the problem ~~which~~ ^{that} precedes the writing of the equation. One of the important goals today of cognitive science and artificial intelligence is to understand the nature of ~~these~~ ^{these} internal representations. I won't try to say in any breadth what we know ^{these matters} about ~~it~~, but ^{I will} give you an example of what I ~~really~~ mean by an internal representation so you can get a gut feeling about it.

We all talk about mental pictures. That gives philosophers a great difficulty, because they tell us that if you have a mental picture, there must be a little man inside looking at the picture, and then what kind of picture does the little man have? And you can get into all sorts of regressions if you allow your mind to stray in that way. So, let's ask what could a mental picture mean? ^{Since you} And I'll ask all of you, you all ^{all} can form mental pictures, I'll ask ~~all~~ ^{such} of you to form a ~~mental~~ picture along with me. Imagine a rectangle, ^{that is} ~~it's~~ twice as broad as it is high, exactly twice as broad as it is high. Now drop a line from the middle of the top of the rectangle to its bottom; ^{it} divides it in two, doesn't it? And the two halves are each squares; you all see the two squares? Now I want you to draw a diagonal from the northwest corner of your rectangle to the southeast corner--from Seattle to Miami. Okay, you've got that line drawn? Now my question is, does that line intersect the line that you drew bisecting the rectangle from top to bottom? ~~How many think it does?~~

~~No more than that? How many think it doesn't, I guess~~ ^{the} ~~is more~~ ~~important question.~~ Now the interesting question here, of course, is not that it does intersect; but I wonder how many of you in the room believe you could prove that it intersects? ^{How many} ~~That you~~ could use the methods of mathematics to prove that it intersects? In fact, that's a rather difficult problem in ^{topology and} geometry. ~~It's so simple that's~~ ^{Although we all instantly and clearly see the intersection,} it's difficult to prove these two lines intersect. If you use analytical geometry, ~~it's a little easier if you~~ ^{to} write down equations for the lines, ^{it's just a little easier,} Do you suppose that you or the homunculi ^u in your head ~~or somebody~~ performed the analytic geometry? Well, maybe.

~~But~~ ^{semantic} the problem of representation is the problem of understanding

how operations of that kind can be performed on information in the human head. And certainly, it appears to us, or it ~~appears~~ to most of us who are researching in this, that those operations are not the operations of deductive logic. They are simple processing operations, but of a ~~very~~ much more, ~~what shall I call it,~~ ^{than logical inference,} a direct kind. I don't have a good label for ~~it~~ ^{them} because there's still a lot of dispute as to how ~~it~~ ^{they} might go on. ~~Now~~ I could carry on my example of that rectangle and add other information, and I could soon convince you that even if you have a picture of ~~that~~ ^{the} rectangle, it's far short of a photograph. ~~I could give~~ ^{If I added} you about three more lines ~~to put and then~~ you would have about as much information as ~~as~~ ^{are able} you ~~were about~~ to keep in that rectangle. ~~So~~ ^{In any case, you can hold} a very limited amount of information, but information that somehow or other is able to generate geometric properties ~~which~~ ^{that} are not explicitly given in the description of the picture. I didn't tell you about that intersection. I never mentioned it.

Conclusion

So, today there are ~~very~~ important open problems in the area of artificial intelligence with respect to how the human head stores and processes these kinds of semantic representations. ~~And again,~~ ^{he talked} Roger Schank yesterday referred to this in other contexts, about other kinds of problems of internal representation. What does this all add up to? What conclusions should we draw from it?

I've already mentioned one of those conclusions. Cognitive science has progressed to the point today where we can begin to understand, not simply toy examples of human intelligence at work, but ~~we can understand~~ ^{also} real examples, ~~that is,~~ examples at the level of school and professional tasks, of intelligence at work. We can ~~very~~ often model that intelligence ~~in the sense of~~ ^{by} building computer simulations ~~of it~~ which track very

closely the human performance and even the human errors. ~~We can also,~~
 If we're not interested in simply modeling the human performance, we can
 also write computer programs which can serve as experts for us. ^{In this way,} ~~We can~~
~~serve to~~ augment human intelligence, and within the past three or four
 years the construction of such experts has become a very popular game
 and even a serious economic enterprise in this country. ~~But~~ as far as
 applications are concerned, ~~it seems to me that~~ in many ways the most
 interesting and the most important ^{possible} are the ~~possibilities of~~ application
 to ~~pedagogy~~. ^{pedagogy} We have operated our institutions of learning for too many
 years on the ~~spring~~ ^{work-spraying} theory that I mentioned earlier--~~not because, well~~
~~yes,~~ because it was the only theory ~~that~~ we had available. Just as
 medicine, when it began to understand the physiology of the human body,
 was able to begin to achieve results in the ~~way of~~ prevention and ~~the~~
 cure of disease that it was not able to achieve simply on the basis of
 folk art and skill, ~~so, I think~~ in the process of education, we are
 going to be able to improve our educational processes at all ~~of the~~
 levels of our schools as we begin to understand what it is that a
 person has to learn in order to engage in an intelligent or skilled
 performance. ~~I~~ think these developments have another implication or set
 of implications that we all sense. ~~That~~ They really go very much to the
 root of the human view of our own species, of ourselves. ~~That~~ Every so
 often there comes along some event in the world of ideas, ^{that} ~~which~~ forces
 humankind to reconsider ourselves, our whole position in the universe, and
 to re-evaluate our sense of worth, ~~and~~ I think the computer is having
 that impact. ^{Copernicus} ~~Copernicus~~ came along, and we could no longer believe that
 our little planet ~~here~~ was at the center of things. We're just cir-
 culating around the sun and the sun isn't even at the center of things.

It's out on the wing of some galaxy and goodness knows where that galaxy is. That was a kind of shocking idea. But we got used to it. I don't think anybody in this room lies ~~awake~~ sleepless at night wondering how to get the earth back into the center of the universe, or being terribly unhappy that it isn't there.

Then came along a man named Darwin. Well, not all of us have gotten over Darwin yet, but most of us have. The great shock of learning that the human species had origins of the same kind as other species, that the mechanisms bringing about its creation ^{were} ~~were~~ the same mechanisms that brought about the creation of other and more humble beasts, ~~in creation~~. We had to give up that idea of our uniqueness after Darwin, and most of us have made our peace with ^{him} ~~it~~. We no longer have to be especially created in that way, ⁱⁿ ~~and~~ that sense, in order to feel ^{our} ~~that~~ human worth.

Today there are people ~~coming along~~, and you've been listening to one of them this morning, who say, ~~and I think this is~~ in the best tradition of ~~biology at least of~~ anti-vitalist biology ^{that} ~~that~~ a human being is a natural organism, ~~is~~ a biological organism. ^{They say that} The reason ~~we~~ human beings can think is because ^{we have} ~~the human being has~~ neurons and other structures in the brain ^{that} ~~which~~ are capable of supporting an information processing system, a symbol system, of the kind I've been describing. This is the reason why we can think.

It turns out that computers, with a completely different kind of ^{from people} hardware, are also able to support information processing, symbol processing, and therefore are also able to think. And so we are beginning to get used to the idea that there are other creatures in creation besides ourselves who are capable of intelligence. Of course

we've known that all along. We know that our dog ^{can} ~~could~~ think and our cat ^{can} ~~could~~ think, but ^{they can} only think little thoughts. We were the people on this earth who could think the big thoughts. And now computers ~~we~~ ^{have} ~~come along~~ ^{come} and from time to time they're thinking some rather big thoughts, indeed. So, we undergo a third challenge to our uniqueness.

Maybe there's a solution to this if it troubles us. Maybe there's ~~a solution to this~~ ^{that} is a solution not only to this specific problem but a solution in a slightly more general sense. Maybe the trouble is not that we're losing any particular sense of our uniqueness, maybe the trouble consists in considering that it is our uniqueness ^{that} ~~which~~ defines us as human beings, our uniqueness ^{that provides} ~~which defines~~ human worth in this world. ^{Maybe the solution is to} ~~And that we can~~ again find a way of relating ourselves to the world in which we ^{live} ~~find ourselves~~--not by finding a new basis for pride, for judging ourselves in some respect ^{that} superior to the rest of nature; Not by setting ourselves apart from nature, but by recognizing at last that we are a part of nature, ^{that} ~~and~~ we must learn to ^{live} ~~leave~~ at peace with nature.