COGNITIVE SCIENCE 4, 33–46 (1980)

# Cognitive Science: The Newest Science of the Artificial*

HERBERT A. SIMON
*Carnegie-Mellon University*

Cognitive science is, of course, not really a new discipline, but a recognition of a fundamental set of common concerns shared by the disciplines of psychology, computer science, linguistics, economics, epistemology, and the social sciences generally. All of these disciplines are concerned with information processing systems, and all of them are concerned with systems that are adaptive—that are what they are from being ground between the nether millstone of their physiology or hardware, as the case may be, and the upper millstone of a complex environment in which they exist. Systems that are adaptive may equally well be described as "artificial," for as environments change, they can be expected to change too, as though they were deliberately designed to fit those environments (as indeed they sometimes are).

The task of empirical science is to discover and verify invariants in the phenomena under study. The artificiality of information processing systems creates a subtle problem in defining empirical invariants in such systems. For observed regularities are very likely invariant only within a limited range of variation in their environments, and any accurate statement of the laws of such systems must contain reference to their relativity to environmental features. It is a common experience in experimental psychology, for example, to discover that we are studying sociology—the effects of the past histories of our subjects—when we think we are studying physiology—the effects of properties of the human nervous system. Similarly, business cycle economists are only now becoming aware of the extent to which the parameters of the system they are studying are dependent on the experiences of a population with economic events over the previous generation.

In artificial sciences, the descriptive and the normative are never far apart. Thus, in economics, the "principle of rationality" is sometimes asserted as a descriptive invariant, sometimes as advice to decision makers. Similarly, in psychology, the processes of adaptation (learning) have always been a central topic, at one time a topic that dominated the whole field of research. Linguistics, too, has suffered its confusions between descriptive and normative attitudes towards its subject. But we must avoid the error, in studying information processing systems, of thinking that the adaptive processes themselves must be invariant; and we must be pre-

*This research was supported by Research Grant MH-07722 from the National Institute of Mental Health, and by a grant from the Alfred E. Sloan Foundation.

pared to face the complexities of regression in the possibility that they themselves may be subject to improvement and adaptation.

It might have been necessary a decade ago to argue for the commonality of the information processes that are employed by such disparate systems as computers and human nervous systems. The evidence for that commonality is now overwhelming, and the remaining questions about the boundaries of cognitive science have more to do with whether there also exist nontrivial commonalities with information processing in genetic systems than with whether men and machines both think.

We are assembled here to take part in the christening of a domain of scientific inquiry that is to be called Cognitive Science. It is not often that a christening is postponed for as many years as this one has been, for even with a conservative reckoning of birthdates, cognitive science has been for some time old enough to vote. There is substantial evidence that the infant was born no later than 1956.

This christening ceremony, then, does not seek to create a new discipline, but to provide a channel for recognizing and handling a set of common concerns among cognitive psychologists, researchers in artificial intelligence, linguists, philosophers, and others seeking to understand the human mind. Understanding the human mind is indeed a venerable goal of cogitation and research. A history of that endeavor must begin no later than with Aristotle.

If that is so, why do I mention the year 1956? That year is important because it signaled a new approach to understanding the human mind, a new scientific paradigm, that today we call the information processing paradigm. In 1956 George A. Miller published an information processing account of the limited capacity of short-term memory (Miller, 1956); Chomsky published one of his first analyses of the formal properties of transformational grammars (Chomsky, 1956); Bruner, Goodnow, and Austin, in their *A Study of Thinking* (1956), introduced strategies as mediating constructs in cognitive theory; and Allen Newell and I published a description of the *Logic Theorist*, the first computer program that solved problems in imitation of humans by heuristic search (Newell & Simon, 1956). A busy year, 1956.

If I may take 1956, then, as the year of the birth of cognitive science—of the analysis of the human mind in terms of information process—the ensuing years have witnessed its steady and moderately rapid growth. The growth is evident whether measured in terms of the research effort going into the field, in terms of the production of new knowledge about mind, or in terms of the acceptance of the information processing approach by the scientific disciplines on which it impinges.

The growth of cognitive science generated new journals in several of these disciplines, when the existing journals were too crowded or too stodgy, or both, to accept the new contributions. *Artificial Intelligence* and *Cognitive Psychology* are just two of these new channels of communication. But they, and others like them, were largely confined to their separate disciplines, and only with the establishment of *Cognitive Science*, about three years ago, was a channel created that cut squarely across the disciplinary boundaries.

## THE STUDY OF INTELLIGENT SYSTEMS

But already my characterization of cognitive science has been too narrow. I have been speaking of the understanding of the human mind as its research goal. Historically, that is perhaps not wholly inaccurate. Until quite recently, the idea of intelligence has always been associated closely with brains and minds, and especially with the human mind. But programs of research in artificial intelligence and in the computer simulation of human thinking have taught us how to construct intelligent systems that are not human, and how to abstract the requisites and earmarks of intelligence from the "hardware" of the brains and electronic boxes that exhibit it.

Hence, I think that most of us today would prefer to define cognitive science as the domain of inquiry that seeks to understand intelligent systems and the nature of intelligence. We have learned that intelligence is not a matter of substance—whether protoplasm or glass and wire—but of the forms that substance takes and the processes it undergoes. At the root of intelligence are symbols, with their denotative power and their susceptibility to manipulation. And symbols can be manufactured of almost anything that can be arranged and patterned and combined. Intelligence is mind implemented by any patternable kind of matter.

I will not elaborate on the topic of symbol structures and their manipulation as the core of intelligence, for that theme will be developed by my colleague, Allen Newell, at a later session of this conference.[1] But in everything that follows, I will simply assume this basis for intelligent behavior, and will turn, instead, to a different aspect of the intelligent systems known to us—their malleability and adaptability, and hence, their fundamental artificiality.

## MALLEABILITY AND ADAPTATION

Intelligent systems exhibit their intelligence by achieving goals (e.g., meeting their needs for survival) in the face of different and changing environments. Intelligent behavior is adaptive, hence must take on strikingly different forms

[1]For an initial statement of this point of view, see Newell and Simon, 1976.

when the environments are correspondingly different. Intelligent systems are what they are from being ground between the nether millstone of their physiology or hardware, which sets inner limits on their adaptation, and the upper millstone of a complex environment, which places demands on them for change.

Systems that are adaptive may equally well be described as "artificial," for as environments change, they must change too, as though they were deliberately designed to fit those environments (as indeed they sometimes are) (Simon, 1969). The task of empirical science is to discover and verify invariants in the phenomena under study. The artificiality of information processing systems creates a subtle problem in defining empirical invariants for them. For observed regularities are very likely invariant only within a limited range of variation in their environments, and any accurate statement of the laws of such systems must contain reference to their relativity to environmental features.

It is common experience in experimental psychology, for example, to discover that we are studying sociology—the effects of the past histories of our subjects—when we think we are studying physiology—the effects of properties of the human nervous system. Similarly, business cycle economists are only now becoming aware of the extent to which the parameters of the system they are studying are dependent on the experiences of a population with economic events over the previous generation (Simon, 1979a).

Finding invariants, then, in artificial phenomena is not an easy task. But that is no counsel of despair for cognitive science. Absolute invariance is very rare in nature, unless it be in the structure of space-time or in the hypothesized elementary particles for which the physicists long so ardently. Biology deals with the laws of systems that have only come into existence in the later stages of the history of our solar system; and if life exists elsewhere in the universe, as we are assured on probabilistic grounds that it must, biologists have no assurance that the basic regularities they observe in life on Earth will also hold for living forms in other galaxies.

What we are searching for are relative invariants: regularities that hold over considerable stretches of time and ranges of systems. What is invariant in adaptive systems will depend on the time intervals during which we observe them. There are at least three time scales of relevance to such systems, corresponding to three different forms of adaptation.

On the shortest time scale, intelligent—hence adaptive—systems continually change their behavior in the course of solving each problem situation they encounter. Hence a prime characteristic of heuristic search, the more so the more successful the search, is that the system gradually takes on the form and behavior that is requisite to adapt it to the environment in which it finds itself.

On a somewhat longer time scale, intelligent systems make adaptations that are preserved and remain available for meeting new situations successfully. They learn. There are many forms that this semi-permanent adaptation can take, and correspondingly many forms of learning. One important form is the accumulation of information in memories and the acquisition of access routes for retrieving it. Learning provides an enormous source for variation in system behavior, hence makes more difficult the search for the elusive invariants.

On the longest time scale, intelligent systems evolve. Their evolution may be Darwinian, by mutation and natural selection in the organismic case. It may equally well be social, through discovery of new knowledge and strategies and their transmission from one system to another. This transmitted inheritance, whether biological or social or both, will also cause a progressive change in system behavior, and will consequently narrow the domain of invariance.

In view of all of these capacities for adaptation, for learning, and for evolutionary change, what room is left for a general science of cognition? What are the invariants we are searching for? We must seek them in the inner and outer environments that bound the adaptive processes. We must ask whether there are any basic characteristics we should expect to be held in common among diverse forms of intelligent "hardware," and we must ask if there are any characteristics that complex problem environments hold in common.

## THE INNER ENVIRONMENT

Our understanding of the invariants associated with the inner environments of intelligent systems is limited to those with which we have experience. From one standpoint, the range of such systems is limited: living organisms and computers. All living organisms make use of essentially the same protoplasmic material, but exhibit a wide variety of organizations. All of the computers that have been built in the course of the short history of that species exhibit remarkable similarity of organization, but have been assembled from a most diverse set of alternative component materials. Hence, over these two classes of intelligent systems, we do encounter a considerable diversity of both organization and material substrate.

In the comments that follow, I will focus on human intelligence and computer intelligence. For some purposes, I will broaden the range to include social insects and human organizations. I do not think the picture would be changed greatly if we looked at other forms of intelligence.

Both classes of systems are symbol systems. They achieve their intelligence by symbolizing external and internal situations and events, and by manipulating those symbols. They all employ about the same basic symbol-manipulating processes. Perhaps that particular invariance arose because computers were made (unintentionally) in the image of man. But no one has yet succeeded in inventing an intelligent system that employs exotic symbol-manipulating processes, and so perhaps the invariance goes deeper than imitation.

None of the systems exhibits a large amount of parallelism in its operation. This assertion is, of course, controversial in the case of human intelligence, and I will not undertake to defend it in detail here (although I firmly believe it to be true and amply confirmed by evidence).[2] In the case of computers, we have apparent counterexamples in such systems as ILLIAC-IV, but the difficulties of programming such parallel computers for all but very special tasks are notorious.

We may conjecture that the real reason for the predominance of seriality in process is that it is very difficult to organize parallel computational systems that require precise coordination of the computations being made simultaneously by the different components. These difficulties defeat human programmers, and apparently they also defeat learning and evolutionary processes.

Where processing is basically serial, all of the relatively labile inputs and outputs of the basic processes can be handled in a working memory of limited size. In the human system, this working memory produces the familiar phenomena of attentional focus. In computer systems, we do not need to be correspondingly limited, unless we wish to be, but roughly similar architectures emerge from the tradeoff between high-speed but costly (hence small) memory components, and low-speed but cheap (hence large) components.

The need for a tradeoff too between flexible adaptation to the environment and coherent attention to goals also seems to point toward mechanisms for attentional focus. Hence, when intelligence is implemented by production systems, a portion of memory is generally designated as working memory or as the "activated" portion of memory in which the condition sides of the productions must be satisfied. We certainly have not learned to design intelligent systems that can take everything (in their memories and in their environments) into account at once; and perhaps nature has not learned to design such systems either. In this case, here is one of the invariants we may seek to characterize and understand.

But am I not ignoring the whole current trend toward multiprocessors, which are the very quintessence of parallelism? And am I not ignoring also those venerable parallel systems, human organizations and ant colonies? I think the apparent contradiction here can be resolved. I spoke above of the difficulty of organizing systems "that require precise coordination of the computations being made simultaneously by the different components." The secret of the human organization and the ant colony is that they do not require coordination of high precision among the individual human employees or the individual ants.

Thought processes can be measured in milliseconds, seconds, or minutes. Coordination of behavior among members of human organizations does not require transmission of information from one to another at millisecond rates, and usually not even with a precision of seconds or minutes. I suppose the limits of such precision are tested in athletic teams, but even in this case the information

[2]From my assertion about the rarity of parallelism, I must exempt, of course, the sensory organs, which clearly are parallel devices. But once feature extraction has been achieved, even recognition processes can be realized readily in real time by serial discrimination nets. What appears to be parallel in central processing (e.g. talking while driving) is almost certainly time sharing in a serial system. See Simon (1979b), Chapter 2.3.

transmitted from one team member to another is very small compared with the rate at which information is processed in each individual head.

Of course if the members of an organization are engaged in independent tasks, or if the needs for coordination are modest, nothing prevents their operating in parallel. All of the numerous examples of parallelism we see in nature seem to conform to this general principle: the rate of inter-component interaction is small compared with the rate of intra-component interaction. Systems having this property are called nearly-decomposable systems, and their near-decomposability has a number of interesting theoretical consequences (Simon, 1969, Chapter 4; Courtois, 1977).

I will predict that as we proceed with the design of ever larger and more complex multiprocessors, their architecture will exhibit with ever greater clarity the property of near decomposability and the quasi-hierarchical structure that goes with it.[3] This will happen independently of whether the designers of these multiprocessors turn to the lessons of human organization (and ant colonies) for design ideas, or whether they ignore the literatures that record this experience and reinvent the wheel.

Perhaps these remarks suffice to give some idea of the nature of the invariants we may hope to discover in the inner environments of intelligent systems. As can be seen from my examples, the invariants are of a relatively abstract kind, and tend to impose constraints on the possible organizations of intelligent systems rather than upon their material substrates. Intelligent systems, it would appear, will necessarily be symbol systems, their high-frequency components will be serial in operation with attentional focus, and the more complex systems will be hierarchic and nearly decomposable.

## THE OUTER ENVIRONMENT

The second source of invariance in intelligent systems derives from common characteristics of the environments to which they must adapt, and in which their behavior takes place. Again, since these environments are of such diverse sorts, the invariants will be of a highly abstract nature.

The environments we are concerned with, those that demand the exercise of intelligence, are problem environments. They do not present obvious paths to the attainment of a system's goals. Frequently, though not always, problem environments contain large, sometimes immense, numbers of alternatives, only a small fraction of which satisfy the goal requirements.

[3] I use hierarchy here in the same sense as in Simon (1969), Chapter 4, not to refer to pyramidal control structure but to describe a modular (nearly decomposable) structure that is "layered" according to the frequencies and temporal precisions of the interactions among modules, among their submodules, and so on. A system with a highly "democratic" control structure can be hierarchic in this sense.

The principal mechanism of intelligence that we have observed (in people or computers) operating in problem environments is heuristic search. The "search" part of the heuristic search process is obvious enough. What are more subtle are the heuristic devices that enable an intelligent system to carry on searches with great selectivity by (1) using information stored in memory to choose more promising over less promising paths, and (2) extracting from the problem environment new information about regularities in its structure that can similarly guide the search. Since we have acquired a considerable knowledge and a considerable literature about heuristic search over the past quarter century, I will not expand upon this topic here (a standard reference is Nilsson, 1971).

Before leaving the topic of the outer environment, however, I should like to call attention to the critical importance of the interface between that environment and the intelligent system: the sensory and motor organs that the latter possesses. This interface presents what is in many ways the most delicate problem in the design of an adaptive system, for it must meet the requirements of both the outer and the inner environments—must, in fact, communicate between them. It is probably no accident that our progress has been much slower in designing effective sensors or effectors, or imitating the corresponding human organs, than in understanding and imitating those intelligent processes that can go on inside the human head or the black box.

## LEARNING

The idea that intelligent systems are highly adaptive and flexible in their behavior could well lead to the notion that their invariants are to be found, not in their behavior or the structures responsible for performance, but in the longer-run mechanisms that bring about the adaptation—their learning mechanisms.

As a matter of fact, learning was a very popular topic in the early history of artificial intelligence research, and an almost dominating topic in cognitive psychology during the period from World War I to the middle 1950s. The historical reasons are not entirely obvious in either case, and we must be careful not to assume that the reason suggested in the previous paragraph was the important one in giving learning its prominence in these two fields.

With respect to artificial intelligence, it is my impression that many workers held the view that it was easier to induce a system to organize itself from scratch, by exposing it to an appropriate sequence of training experiences, than it was to provide it with the knowledge it would need for expert performance. There was perhaps also a borrowing of attitudes from psychology, where learning was then viewed as the core of the subject, and where Hebb and others were considering how nerve nets might organize themselves.[4] Finally, there was prob-

[4]For an important early example of work developing this point of view, see Rochester, Holland, Haibt, and Duda, 1956; and for a basic critique of the approach, see Minsky, 1963, Section III.

ably the desire to avoid the charge, commonly directed toward artificial intelligence systems, that the intelligence was really in the programmer, and not in the system at all. If the programmer only provided a potential for learning, and not the program for the finished performance, then this accusation could not stand.

Whatever the reason, many investigators in the early years followed the learning route, and I think it fair to say that they largely failed. All of the expert systems that have been produced up to the present time have been given directly all or most of the knowledge and the problem-solving strategies on which their expertness rests.

We should be careful not to extrapolate the research experience of the past two decades to the research programs for the next two. Over the past five years there have been many signs of a revival of the learning enterprise, but with a viewpoint quite different from the original nerve net or simple reinforcement paradigms. Systems like METADENDRAL (Buchanan & Mitchell, 1977), which induce their own theories, or like Lenat's AM, which acquire new concepts that help them, in turn, to discover still others, are certainly learning systems. So is UNDERSTAND (Hayes & Simon, 1974), which generates, and then uses, problem representations; Langley's (1979) BACON, that induces scientific laws from data using recursive procedures; Neves' (1978) program that learns skills by analysing worked-out textbook examples; and many others.

What is characteristic of the new generation of learning programs is that most of them are basically problem-solving programs, capable of undertaking heuristic search with a reasonable armory of methods like generate-and-test and means-ends analysis. They do not start out in anything like the barebones fashion of the earlier generation of self-organizing systems.

Let me now turn back to psychology's preoccupation with learning during the period that just preceded the information processing revolution. I suggested earlier that the motivation for this preoccupation might have been the concern that only learning parameters could be invariants in an artificial system. A review of the writings of the most prominent investigators does not support this hypothesis. I can find only a weak hint of it in the first edition of Hilgard's *Theories of Learning* (1948), and no hints at all in the books of Hull, Watson, or Thorndike.

The historical reasons appear to be quite different. In the first place, there was an applied motivation. An understanding of learning was key to an understanding of the educational process. This motivation was quite clear in the work of John Dewey, of Thorndike and others. But even stronger than the urge of relevance to education was a philosophical goal that ties the learning research back to the earliest concerns of psychology.

Psychology had its birth, of course, in philosophy. And the philosophical question that led most directly to a psychological one was the epistemological question: How can we know the world outside ourselves, and how can the mind store knowledge about that world? It is a quite easy and direct step from asking how we know the world to asking how we acquire that knowledge—how we learn (as well as the closely related question of how we perceive). It would seem plausible that this epistemological concern accounts for psychology's preoccupation with learning during a period when psychology was relatively unconcerned about application, sometimes almost belligerently so.

I would now like to leave these historical questions and point to what almost amounts to a contradiction in the return to an interest in learning that I asserted has been taking place in the past five years. If most of our progress in understanding expert skills (both in psychology and artificial intelligence) has been gained by studying and constructing expert performance systems, why are we now turning back to an "empty box" approach?

Part of the answer has already been suggested. The new learning systems do not much resemble those of two decades ago. They start out with a much more complex and sophisticated system than did the earlier nerve-net and perception approaches. Therefore, we have reason to hope that they can climb higher on the ladder of skill, and we already have experience that supports that hope.

A second part of the answer is that we now have a vastly better picture than we had earlier of the goal of the learning: of the expert performance systems that the learner is seeking to attain. We know a great deal about how expert knowledge is stored in associative, list-structure memories, and how this knowledge can be organized for effective access.

A third part of the answer is that most of our expert performance systems are now constructed as production systems, and we have clearer ideas of the mechanisms needed to enable a production system to bootstrap itself (adaptive production systems (Waterman, 1975)) than we ever did for systems organized as hierarchic structures of closed subroutines.

I am not predicting that we are soon going to abandon our current emphasis, in psychology and artificial intelligence, on understanding performance systems in favor of studying learning. Much less am I urging that we do this. But as the manpower available for research in cognitive science continues to grow, there probably are hands and heads enough to go forward on both fronts.

In our study of performance, however, we must not imagine invariants where there are none. Since intelligent systems are programmable, we must expect to find different systems (even of the same species) using quite different strategies to perform the same task. I am not aware that any theorems have been proved about the uniqueness of good, or even best, strategies. Thus, we must expect to find strategy differences not only between systems at different skill levels, but even between experts.

Hence, research on the performance of adaptive systems must take on a taxonomic, and even a sociological aspect. We have a great deal to learn about the variety of strategies, and we should neither disdain nor shirk the painstaking, sometimes pedestrian, tasks of describing that variety. That substrate of description is as necessary to us as the taxonomic substrate has been to modern biology. Within the domains of cognitive science, perhaps only the linguists (and to some extent, the developmental psychologists) have had a tradition of detailed description instead of a tradition of experimentation in search of generally valid truths.

As for the sociological aspect, performance programs are in considerable part the product of social learning, and we must not expect the performance programs of the twentieth century to be identical with those of the nineteenth, or the tenth, or the first. The prospect of far-ranging changes in human cognitive programs again reintroduces important philosophical problems that have received only modest attention in psychological research. In what respects is the mind of the Greek citizen the same as or different from the mind of modern man? Is thinking in an African village the same process as thinking on a mechanized farm? Experimental material is available for studying the latter kind of question, if not the former, and a little such research has been done. But as we begin to take more seriously the proposition that human cognitive programs are determined as much by social and historical forces as by neurology, our research emphases will surely shift.

## NORMATIVE AND POSITIVE SCIENCES

In artificial sciences, the positive (descriptive) and the normative are never far apart. Thus, in economics, the "principle of rationality" is sometimes asserted as a descriptive invariant, sometimes as advice to decision makers. The business firm is assumed to act so as to maximize its profits, and theories of how the economy works are erected on this assumption of maximization. On the other hand, specialists in operations research, bringing to bear the sharp tools of linear programming, integer programming, queuing theory, and the like, offer to help business firms maximize profits where they are not so doing.

Similarly, in psychology, the view that intelligent systems are adaptive, can learn, and evolve does not prevent us from studying them in order to discover how to improve their learning or problem-solving powers. The feeling of contradiction between the positive and normative views is less acute in psychology than in economics precisely because in the former field we only rarely assume that adaptation or learning or evolution maximizes something and therefore is not open to improvement.

The recent emergence of sociobiology, with its claim that fitness arguments can be used to explain, and even predict, how things are, threatens to introduce new confusion in the relation between the positive and the normative. I will have more to say about that in the next section of this paper.

Linguistics, too, has suffered its confusions between descriptive and normative attitudes toward its subject. I do not refer so much to the old-fashioned dictionary maker's notion that his task was to prescribe "correct" usage, but to the more modern idea of a language sharply defined in terms of "competence" and "grammaticality." There is a continuing danger that focus upon an ideal competence that resides in some kind of Platonic heaven (or a Cartesian one) will impose normative constraints on the study of actual language behavior.

I am not arguing against the desirability of normative science in any of these domains, but simply against a confusion of the normative with the positive—of what ought to be with what is. Artificial intelligence, in particular, certainly addresses itself to normative goals, and ought to. It is interested, in its applied aspects, not only in understanding intelligence but in improving it. Perhaps we need to recognize this explicitly by speaking of cognitive engineering as well as cognitive science. If we do, however, I hope that the two ventures will keep in the closest relation with each other, as they have done through the past quarter century. The dangers of confusing the normative with the positive are slight compared with the losses that would be suffered from isolating the science from its engineering applications.

## EVOLUTION AND OPTIMIZATION

I want to return now to the topic of evolutionary theory, especially in its contemporary incarnation as sociobiology (Wilson, 1975). Economists, in particular, are seizing upon evolutionary arguments to expand their domain from strictly economic phenomena to the whole domain of individual and social behavior (Simon, 1979a).

The slogan, of course, of evolutionary theory is "survival of the fittest," certainly a claim of optimization. It is important to examine what this means, and can mean, in the world as we know it. "Survival of the fittest" refers to the outcome of the competition among species seeking to occupy particular ecological niches. In the simplest form of the theory, there can be at least as many surviving, fit species as there are niches—that is, an enormous number. Moreover, the niches themselves are not determined by some inflexible, invariant environment, but are defined in considerable measure by the whole constellation of organisms themselves. There can be no lice without hairy heads for them to inhabit, nor animals without plants. Hence, it is not obvious what optimization problem, if any, is being solved by the process of evolution. At most, the occupancy of each niche is being locally "optimized" relative to the entire configuration of niches.

Two formal developments, the theory of games and linear programming theory, give us important insights into the difficulties of using optimization assumptions to predict the behavior of complex systems. From the theory of games we learn that in competitive situations it may be impossible to define what we mean by an optimum in an unambiguous way, much less to guarantee that unique optimal solutions of the game exist. From linear programming, we learn that very strong conditions must be met by a system (e.g., linearity of the criterion function, convexity of the feasibility space) to guarantee that local maxima will be global maxima, and that when these conditions are not met, it is very difficult—in fact, usually impossible—to devise algorithms that will discover global optima within the limits of a reasonable computational effort.

For these reasons, cognitive science is likely to remain a science of systems that satisfice—that find tolerable solutions for their problems—rather than a science of systems that optomize—that adapt perfectly to their environments. But to understand satisficing intelligence, it is necessary to understand the process through which that intelligence is exercised. It is not enough simply to infer what the intelligent behavior "must" be from the description of the environment and the conditions of optimization. But I will not elaborate this argument further, for I expect that most cognitive scientists in this audience would not disagree with it.

## CONCLUSION

It might have been necessary a decade ago to argue for the commonality of the information processes that are employed by such disparate systems as computers and human nervous systems. The evidence for that commonality is now overwhelming, and the remaining questions about the boundaries of cognitive science have more to do with whether there also exist nontrivial commonalities with information processing in genetic systems than with whether men and machines both think. Wherever the boundary is drawn, there exists today a science of intelligent systems that extends beyond the limits of any single species.

Intelligence is closely related with adaptivity—with problem-solving, learning, and evolution. A science of intelligent systems has to be a science of adaptive systems, with all that entails for the difficulty of finding genuine invariants. Cognitive science is a science of the artificial. It is concerned with phenomena that could be otherwise than they are—and which will be altered continually as they adapt to the demands of their environments.

So long as we do not confuse adaptability with the ability to attain optimal solutions, cognitive science will be a basically empirical science. Inference from optimality conditions will play a modest role in helping us to discover how things are. Most of our knowledge will have to come from extensive painstaking observation of the vast variety of intelligent systems that exist in the world, the equally vast variety of programs that these systems acquire and employ, and from formal theories—mainly in the form of computer programs—induced from that body of observation.

## REFERENCES


Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking*. New York: Wiley, 1956.

Buchanan, B. G., & Mitchell, T. M. Model-directed learning of production rules. Presented at the Workshop on Pattern-directed inference Systems, Honolulu, Hawaii, May 1977.

Chomsky, N. Three models of the description of language. Proceedings of a Symposium on Information Theory. *IRE Transactions on Information Theory*, Sept. 1956, IT-2(*3*), 113–124.

Courtois, P. J. *Decomposability: Queueing and computer system applications.* New York: Academic Press, 1977.

Hayes, J. R., & Simon, H. A. Understanding written problem instructions. In L. W. Gregg (Ed.), *Knowledge and cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1974.

Hilgard, E., & Bower, G. H. *Theories of learning* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall, 1975.

Langley, P. Rediscovering physics with BACON.3. *Proceedings of the 6th IJCAI–Tokyo, 1979, 1*, 505–507.

Miller, G. A. The magical number seven. *Psychological Review*, 1956, *63*, 81.

Minsky, M. Steps toward artificial intelligence. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought.* New York: McGraw-Hill, 1963.

Neves, D. A computer program that learns algebraic procedures by examining examples and by working test problems in a textbook. *Proceedings of the 2nd CSCSI Conference*—Toronto, 1978, 191–195.

Newell, A., & Simon, H. A. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, March 1976, *19*, 111–126.

Newell, A., & Simon, H. A. The logic theory machine. *IRE Transactions on Information Theory*, Sept. 1956, IT-2(*3*), 61–79.

Nilsson, N. J. *Problem solving methods in artificial intelligence.* New York: McGraw-Hill, 1971.

Rochester, N., Holland, J. H., Haibt, L. H., & Duda, W. L. Test on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on Information Theory*, 1956, IT-2(*3*), 80–93.

Simon, H. A. Rational decision making in business organizations. *American Economic Review*, Sept. 1979, *69*, 493–513.

Simon, H. A. *Models of thought.* New Haven: Yale University Press, 1979b.

Simon, H. A. *The sciences of the artificial.* Cambridge, MA: M.I.T. Press, 1969.

Waterman, D. A. Adaptive production systems. *Proceedings of the 4th IJCAI*—Tbilisi, Georgia, USSR, 1975, *1*:296–303.

Wilson, E. O. *Sociobiology: The new synthesis.* Cambridge, MA: Belknap Press of Harvard University, 1975.