standing problems concerning the complexity of motion planning and finds a collision-free path for a jointed robot in the presence of obstacles. Canny's new algorithm for this "generalized movers' problem," has a single exponential running time, and is polynomial for any given robot. In deriving the single exponential bound, Canny introduces two powerful tools; the generalized (multivariable) resultant for a system of polynomials and Whitney's notion of stratified sets. He has also developed a novel representation of object orientation based on unnormalized quaternions. After dealing with the movers' problem, Canny derives several lower bounds on extensions of the problem: finding the shortest path among polyhedral obstacles, planning with velocity limits, and complaint motion planning with uncertainty. He introduces a clever technique, "path encoding," that allows a proof of NP-hardness for the first two problems and then shows that the general form of compliant motion planning is nondeterministic exponential time hard.

A ROBOT PING-PONG PLAYER
Experiment in Real-Time Intelligent Control

Russell L. Andersson

ISBN 01101-8, March 1988, 300 pp., $35.00,
15 minute tape shows the player in action
ISBN 01105-0, VHS: $55.00;
ISBN 01106-9, U-MATIC: $60.00
The MIT Press

This tour de force in experimental robotics paves the way toward understanding dynamic environments in vision and robotics. Anderson introduces the first robot able to play, and even beat, human ping-pong players.

KNOWLEDGE-BASED SYSTEMS FOR MANAGEMENT DECISIONS

Robert J. Mockler

ISBN 516906, April 1988, Cloth $32.00
Prentice-Hall
Englewood Cliffs, NJ 07632

Discover how to use AI tools, especially expert system development shells, and create knowledge-based systems for managerial decision making.

Here is an emphasis on structuring the decision situation in a way that makes putting these systems onto the computer easier. Now, non-technical managers can create actual knowledge-based systems that work with no prior knowledge of computers necessary to develop them.

KNOWLEDGE-BASED SYSTEMS FOR STRATEGIC PLANNING

Robert J. Mockler

ISBN 516914, April 1988, Cloth, $32.00
Prentice-Hall

Here's an opportunity to use AI technology in creating knowledge-based systems for strategic planning. Innovative features introduce readers to AI and expert systems, structured situation models, scenario development, dependency diagrams (complete graphic pictures of prototype systems), and guidelines for putting the system onto computers using expert system shells. This book includes detailed and abundant examples of actual working knowledge-based systems with tutorials provided for using major expert system shells.

ARTICLES

Winter 1988 Daedalus

Joseph Agassi
D69@TAUNOS.BITNET
Tel-Aviv University and York University, Toronto

The last issue of Daedalus, Journal of the American Academy of Arts and Sciences, Winter 1988, is devoted to artificial intelligence--AI, for short. It is always interesting to know what is the official position of the intellectual establishment on intellectual matters, especially where much grant allocation is concentrated.

The preface, by Stephen R. Graubard, Editor of The Academy and of Daedalus, opens with an admission that sets the tone for the whole issue: The label of artificial intelligence has helped create a myth: a duplicate of the human intelligence is made in the computer lab! It was "a kind of hubris, ... unbecoming and unnecessary."

That is all. The second paragraph begins with "Yet." How should one read this "yet," this move to the positive side of a balance sheet, once the negative side is swiftly done with by the admission that the term AI is an unbecoming and unnecessary kind of hubris? What is the balance going to look like? Will it illustrate yet again the famous fact that every cloud has a silver lining or will it present one of these rare cases in which the water wasted is miraculously returned to our bottles as wine? For, certainly a lot of grant money was spent on AI. Was that waste or intelligent investment? We do not know. The matter can stand an investigation.

Background information. In a collection of essays, especially one on technical matters geared to the non-specialist, a high degree of redundancy is inevitable. Naturally, more systematic and detailed treatments of background material are available. The issue at hand is the present state of the art, and the inner dispute; an outline of the necessary background to this is offered here first.

The definition of AI, not surprisingly, is already a bias. In the most biased--initial--behaviorist definition, the inner life of an intelligent being is ignored and AI is viewed as any successful emulation of intelligent behavior, where success is defined as the emulation that fools the expert (Turing's test, 1950). A little less biased is the idea that any being is intelligent if it (would pass Turing's test showing that it) can (1) learn a natural language and (2) create art and science; AI, then, is the program to create an algorism that can make a machine do these things--or at least a theory of it (1956-61).

Computer science and computer technology are terribly clever and exciting and they have doubtlessly fruitfully interacted two-way with many fields of study, including philosophy, logic, mathematics, psychology, and neurophysiology. AI enters the picture only when these exciting developments help emulate intelligent conduct: the interdisciplinary work is but a preliminary to that. This is a point all too often overlooked and it creates unpleasant impressions. To take an example, the expert-

systems programs now in the market for sale are computer programs which to a significant extent may help a specialist the way an expert may. Indeed, expert systems are created by teams of computer experts, each with the help of experts from any one other field For example, a diagnostic expert system for heart specialists has been shown to spot digitalis poisoning in a patient faster than the average heart-specialist. Expert systems, thus, are naturally taken by defenders of AI as vindicating its programs by their success--admittedly partial, yet quite significant. This argument invites the unjust ridicule of expert systems as mere [!] computerised dictionaries of sorts (see pp.148, 215,250-1 and 270; see, however, p.78), since at issue is the intelligence of the program not of the programming team.

When does the program rather than the programmer exhibit intelligence? When it beats its maker in a game of chess? The fact of the matter is that we do not know. As Yehoshua Bar-Hillel has forcefully argued, one can beat a chess program, however intelligent, by making a move so stupid that it has been overlooked by its maker: the program cannot distinguish the stupid from the clever. This is a matter of a great dispute. A minute aspect of it will come up later on, in the discussion concerning the lack of commonsense flexibility, the so-called brittleness, of most computer programs (pp.149, 196-7).

From the definition to the history of the interactions of computer science with other, AI-related fields.

Classical associationist psychology described the perception of objects as the association of perceptions of elementary items (known as sense data) and memory as the residue these leave, like grooves left on the hard soil by passing vehicles, provided the same route is passed repeatedly. This theory was never any good; it was refuted better than any other theory ever was; it is still alive and kicking; for example, it still backs the disastrous practice of learning by rote; it still animates much of current research--in AI, philosophy, psychology, education. Worst of all, it is often taken for granted.

Neurons were discovered in 1940; in 1947 D.O. Hebb declared neural paths to be the putative paths-and-grooves of memory. Yet by then computers with memory banks were already common knowledge and computer memory was more like the written page than like scratches repeated on a hard surface. The theories making analogies with computers therefore simply had to be heretical and break away from associationism. All computer experts know that the problem with memory is both retention and retrieval, yet despite Plato and Freud, who viewed only retrieval as problematic, most writers on memory speak of its problems as those of retention--even in this book! (See p.114, and cf.p.151.)

Already in the earliest days of computers, in 1943, Warren S. McCulloch and Walter H. Pitts had introduced a formal neural network: a network of abstract neurons each operating one of the simple logical operations of conjunction, disjunction or negation, just as in computers. In 1956 a historic meeting took place in Dartmouth College at the invitation of John McCarthy, where the concept of AI was introduced and where Allen Newell and Herbert Simon presented there a computer proof of some (trivial) logical theorems. Marvin Minsky and Seymour Papert met there and then cooperated on constructing AI systems In 1958 Frank Rosenblatt described a complex of formal neurons, a perceptron, which can learn by trial and error.

This was a quiet revolution away from associationism In 1961 Minsky presented the program for AI research as algorizing effectively anything recognizable as intelligent, in which the associationist bias is still manifest. Soon a model of associations was created as a complex of perceptrons--with associations as composite trial-and-error processes, however. Studies of individual brain cell functions began in the late 60's by David Marr and others, who offered hypotheses dividing the logical functions of the quasi-associationist formal model of neural networks between different kinds of cortical cells. Minsky and Papert showed in 1969 that the program of researches presenting the brain as a set of perceptrons is hopeless. Their objection was met in the 80's by a modification of the original program, now known as the connectionist program. The new program is a break-away from associationism into the terra incognita of systemism. (Terminology is still unsettled; the position between mechanism and classical holism is labelled by W.V. Quine modified holism and by Mario Bunge systemism.) The leading new connectionist essays were assembled in a best selling volume, Parallel Distributed Processing (where the parallel processing is the cooperation of many units, and where what is distributed, namely, not localized like in holographic memory are the memory and programs), 1986. The present volume is a follow-up by friend and foe.

The book's structure. Its 310 pages contain 14 items. The first item sets the issue within the AI community and the second sets the historical and philosophical background to it. Two items then explore the very concept of AI and five discuss the issue at hand. The next item, its authors claim, transcends the issue--thus making the rest of this volume obsolete, perhaps. Then comes something that is indisputably substandard, nicely leading to an attack on AI as humbug, followed by a counter-attack. The close is an overview by the initiator of AI. The structure could be improved upon by a more energetic editor.

Here then is the summary; comments are in square brackets.

1. Seymour Papert, professor of media technology and director of the Learning and Epistemology Group at MIT, "One AI or Many?"

There are two schools of AI thought [not many], the old-style programmers who emulate brain processes on the computer, and the new style connectionists, who study brain physiology. Each claims full success for itself in the very near future. The author himself is no party to the dispute, as it is based on a category mistake rooted in "the quest for universality of mechanism" (p.7): each can try to be as universal as possible without limiting the other, as they operate on different levels. [This discussion is impeccable. Were the paper terminated here, it would raise the question, what was the dispute in the first place? Yet we are in the middle of the paper: as this question is delicate, answering it takes a few pages.]

The conflict between the two AI schools, was [not intellectual but] financial: they were quarreling over grant moneys (p.7). The connectionists' program has been refuted by Minsky and Papert; now the connectionists hope to achieve great success in no time with a new program to employ parallel distributed processing. This new program is a waste of time and of scarce grants moneys. Their appeal is but the appeal to the catch phrase "parallel distributed processing."

[What is the difference between the two schools now? why does the author think the new-style connectionist program hopeless? what has become of his view that there is room for the old and the new? Does he belong to

the old-style school of the programmers or is he neutral? The paper begins with a plea for pluralism and ends by condemning one of the two schools extant! As to his complaint that the opponents are using a catch phrase to secure more grant moneys, it comes after the editor's admission that all AI people do that!]

2. The Dreyfus brothers, Hubert L., the philosopher, and Stuart E., the engineer, "Making a Mind Versus Modelling the Brain: Artificial Intelligence Back at the Branching Point."

This is a history of the dispute since its beginning in the mid-fifties, as stemming from the older, traditional, philosophical one, between the mechanist-reductionists and the anti-reductionist holists. [It is difficult to know what the disagreement is about when merely the contributions of two philosophical schools of thought are presented--especially since no single empirical study is ever exclusively confined to the ideas of one school. With scarcely any communication across school lines (both in philosophy and in AI studies), an unschooled reader will despair. As the Dreyfus brothers view the messages of the two philosophical schools as almost identical, understanding them is hard even for the philosophically adept. Their identifying systemism with holism is a bias.]

The question, however, is, what is the AI disagreement? The answer remains [vague to the last]: it is between competing research programs, the mechanistic-reductionist and the holist. [The rest of the Dreyfus paper is not clear; catch phrases are of no help here: catch phrases transfer allegiance from one camp to another too easily unless prevented by statements that certain definite contentions are characteristic of one side and rejected by the other.] Remarks like, "Minsky and Papert were so intent on eliminating all competition ... while completely ignoring ... " (p.22) give the tone to the rest of the paper. [They tacitly validate Papert's complaints. They sound as if research without grants is impossible:] "... was discredited along with hundreds of ... research groups ... research money dried up ... had trouble getting his work published ..." (p.24). [When publishing counts in competition, it gets hard to publish. Publishing should be geared to readership, not to grantsmanship. What should be done about this? No answer. Pity.]

3. Robert Sokolowski, philosopher, "Natural and Artificial Intelligence."

A printed page is already both artificial and somehow intelligent: it is readable; yet as truly intelligent reading is intentional, the printed page is not; nor is the computer. [This is indisputable but does not impinge on the dispute at hand; nor does the author report the AI response.]

4. Pamela McCorduck, author, "Artificial Intelligence: an Aperu."

A readable text is intelligent; so is a program generating or transforming it. It may then be a piece of computer art or an expert systems program. Much can be learned about our notions of art [and of expertise]. [This depends on the details of programs. In general, whatever can be formally described can be reproduced by rote and should be delegated to machines. Art is what at the time goes beyond that. Also, this paper differs from its predecessor; their disagreement is unstated. This is poor editing.]

5. Jack D. Cowan and David H. Sharp, mathematical biologist and theoretical physicist, "Neural Nets and Artificial Intelligence."

The history of the development of artificial neural nets. Final section: "There is still a very long way to go before any kind of truly intelligent robot can be produced" (p.114). Technical problems aside, how can one ascribe intentionality of the computer? and what is learning? "It is hard-wiring that embodies prior knowledge and, in a sense, the intent of the designer. ... In a sense, evolution has acted not as a trainer to soft-wire neural nets but as a critic to hard-wire them ... [This is the neo-neo-Lamarckism of Schrdinger's What is Life? which permeates this book.] Should we be expected to telescope billion years of evolution ... into a few decades of neural net and AI research ... ? Until we understand how ideas and intentions are embodied in the human brain, rapid progress is unlikely. On the other hand, developments ... We predict that the top-down approach of conventional AI and the bottom-up approach of neoconnectionism will eventually join to produce real progress in ... experimental epistemology, the study of how knowledge is embodied in brains and may be embodied in machines."

[As there is no clear division between hard and soft wiring in computers, it is hard to assess all this. Not is it true that the difference between the two schools is methodological (top-down being hypothetico -deductive and bottom up being inductive generalizations from facts). But the disclaimer may be true concerning difference of opinion between old-style and new-style AI--there may be none except at most as to methods and the order of priority of investment of effort.]

6. Jacob T. Schwartz, mathematician, "The New Connectionism: Developing Relationships Between Neuroscience and Artificial Intelligence."

Unlike computers, neurons work in parallel and each of them has many non-localized functions. The clue to brain function theory is that "mental (especially sensory) processes seem to be of very restricted "depth," in the sense that not many successive elementary neural reactions are required to form the higher level reactions that the brain generates. There is simply no time..." This, is admittedly not much of a clue, given that a switch may be simple or as complex as one wants (within the capacity of the machine in question). [This makes associationism pass.] One more detailed clue is the way visual and tactile sensations are mapped in the brain: at first retained in a simple geometric image [the author exhibits an associationist tendency, but treads softly] and then somehow transformed. Another clue is from embryology. In the embryo most cells can move; brain cell send potential synapses instead. The random manner in which this occurs suggests a high degree of non-specificity, though the (morphological, biochemical, and other) differences between the kinds of neurons in the brain, and even their numbers, suggest specificities in need of study. In any case, "biological systems are not wired precisely enough to support this extremely delicate style of information processing" that computers possess (p. 132). The theory of the cerebellum as a set of simple mechanisms capable of learning by conditioned reflex is of help, but "learning-based theories of the origin of neural functions" are still hardly of any use: "we know hardly anything yet about the actual locus or mechanism of other memory storage within the brain and even less about the way memories are modified to accomplish abstract learning" (p.134). [That conditioned reflex is still considered attractive even while its poverty is admitted is fascinating--quite apart from the fact that the conditioned reflex model in question is supposed to be of the most advanced part of the human brain, and quite apart from the fact that the deviations from the conditioned reflex here noted are in the selection and the modification of information.] The very

basis to the whole project, the ideas of thresholds, excitation and inhibition, should undergo a critical reexamination (p.136).

[The meat of the paper is in its conclusion:] Analogy between computer and brains is sheer conjecture (p.137). Computer simulation may help neuroscience, neuroscience can hardly be expected to help design better computers (p.136). There is an exception to this already: analog computers are less stable and accurate than digital ones, but more brain-like and so may be preferable for "the processing of streams of incoming sensory information like audio information or moving images" (p. 139). "Consequently, there is reason to hope that analog networks can process sensory data in a manner that will profit from ..." [There is trouble with the quotation here. Its logic should demand that it should go thus: there is hope that computer data processing research will profit from studies of neuroscience;] the quotation gets somehow lost in detail (p.139). The last two pages express hope that the two branches of AI will one day unite. [This clearly implies that computer simulation is the same as AI! Possibly; but being contested elsewhere in this book it reflects poor editing.]

7. George N. Reeke, Jr., and Gerald M. Edelman, both biologists, "Real Brains and Artificial Intelligence."

AI is as artificial as the dentistry of Aristotle who never bothered to look at Mrs. Aristotle's mouth to check his claim that women have less teeth than men. [Perhaps he was misled by one case!] That the goals of AI and neuroscience are similar has been obscured by two errors: certain epistemological conclusions from the views of Turing and Church of computer as universal problem-solving machines (namely, the taking of idealized cases as if they are real, p. 148) and the view of the brain as a collection of units which exchange chemical signals (namely, the mechanistic view rather than systemic one). "As biologists seeking to understand the nearly dogmatic neglect" of biology by old-style AI researchers, they ask, what are the AI researchers' goals and methods (p.144)?

The success of old-style AI, whatever AI is, the successful application of physics and engineering to computers, rests on taking as basic and unanalyzable some categories and some information about them; this leads to the the proposal to consider perception and intellectual processes as algoristic. [Clearly the authors identify old-style AI of the programmers with programming, though it is the view that all intelligence is programmable. Editor!] This is not evolutionary as it blocks any attempt to explain the rise of characteristics--such as categories in the brain--as adaptive mechanisms. This way biology "might contribute to further progress in AI"(p.145).

[The book's middle thus meets its declared point!] In 1961 Marvin Minsky outlined the old-style AI program as that of searching for effective procedures for search, pattern recognition, planning and induction. This includes some objectionable "epistemological assumptions" (p.146). [Of course: effective procedures for scientific progress is a traditional dream (inductivism). If there are such effective procedures and Professor Minsky will find them, then he will already have made a tremendous contribution to knowledge. When it will be implemented, there will be no more need for human research! And, of course, the truth about categories and information will then not for long remain hidden! The traditional view of scientific research as algoristic is not in the least evolutionary. (The same holds for the theories of associations and conditioned reflexes and so on: their enormous attraction despite their obvious faults is thus explained as the attraction of inductivism as rooted in the dislike for responsibility.) This has been observed for many times in the last century by many authors; chief among them is, perhaps, Sir Karl Popper. Clearly, here the authors have won a victory before this battle has started. But how is that to effect AI research remains to be seen.] The old-style programmers' assumptions are limited; their algorisms are not designed to take care of the limitations involved and are therefore "brittle". when they meet their limits they go over them and "crack." The best solution to the problem of brittleness is only a less brittle program; AI is thus a mere ideal and not the best (p.149ff). [See end of this review.]

Parallel computation is so complex that the only way to learn what a programmed parallel machine can do is let it run, so that each machine is unique and thus not really programmable (p.152). Current computer models of the brain contain too many specific unrealistic assumptions (p.153). In general the computer is taken as passive. The programmer determines in advance the code to feed it information with, the categories which it should deal with and some procedures (p.154). Computers thus cannot adapt (p.155). [This smacks of Lamarckism, as much of the book does.] In real live neural systems there is no strict heirarchy, no single neuron is indispensable with, there is an enormous diversity of kinds of neurons, and more so in more evolved species. "Only patterns of response over many neurons can have functional significance"(p.156), responses depending not on accuracy, speed or efficiency, but on overlapping wide-range "repertoires" of functions. These are more suitable for unprogrammed systems in unknown hostile environment. The authors have constructed automata, "selective recognition systems" which "address some of the problems of the standard AI paradigm by avoiding preestablished categories and programming altogether" (p.161): each computer in the system is programmed to simulate a neuron, and is told nothing about functions or about programs. One resultant automaton could "act upon the environment to form a complete autonomous behavior" (p.161).

[Here is the place to turn back to Papert's critique (p.11). "Although its models use biological metaphors, they do not depend on technical findings in biology any more then they do on modern supercomputers." This is too little for comfort. This is not to endorse the contentions here quoted but to note the editor's neglect: the point, correct or not, demands a better comment from the opponent, especially form one who views the new-style work as mere bogus. The authors understandably describe their automata too s to permit assessment, and it seems that the right way to go about it is to scrutinize their claims in diverse ways. Do the parallel computers really start with no program or are they able to modify apriori given ones? Is the way a computer learns to categorize not also by modifications? Papert's essay is of no help here, both because of his cavalier attitude and because of his associationism.]

8. W. Daniel Hillis, the inventor of the connectionist machine (as his MIT doctoral project!), "Intelligence as an Emergence Behavior; or, The Song of Eden."

Neither emergence nor intelligence are understood, yet the idea of an emergent intelligence is attractive as a possibility of "constructing intelligence without first understanding it" thus not giving way to mechanistic reductionism The shift from sequential to parallel processors "is not a deep philosophical shift, but it is of great prac-

tical importance, since it is now possible to study large emergent systems experimentally" (p.176). [This is very nice.]

Example. A preverbal proto-human race develops songs by mimicry. Songs are parasites on the singers. They survive by the specialization of the moods they express: the survival value is in their usefulness to the community of singers as means of communication and thus as levers for intelligence. [It is time to notice the enormous fruitfulness of Samuel Butler's "Darwin among the machines"!] This raises the question of the size of the storage of the brains of the proto-humans. Living memory is distributive--non-localized, as in a hologram--and so its size is hard to assess. The author assess our storage capacity as surprisingly small. He similarly finds it hard to assess the importance of sensory-motor functions for the growth of intelligence; some such nexus is undoubtable.

Little understanding is needed for construction. This is why constructing artificial intelligence and emergent systems, including emergent artificial intelligence, is so challenging right now. "I have recently been using an evolutionary simulation to evolve programs to sort numbers. In this system, the genetic material of each simulated individual is interpreted as a program specifying a pattern of comparisons and exchanges. The probability of an individual survival in the system is dependent on the efficacy and the accuracy of this program in sorting numbers. Surviving individuals produce offspring by sexual combination of their genetic material with occasional random mutations. After tens of thousands of generations a population of hundreds of thousands of such individuals will evolve very efficient programs for sorting. Although I wrote the program for the simulation that produces the program, I do not understand the detail....If the simulation had not produced working programs, I would have had very little idea about how to fix it" (p.188.). "The result would be not so much an artificial intelligence, but rather a human intelligence sustained within an artificial mind." "Of course, I understand that this is just a dream..." (p.189).

[This is thought-provoking. The reification of songs and other things is challenging. Is a perceived bit of information a thing? Is this a metaphorical use of biology? The author does use questionable analogies, such as his view as of evolutionary value the survival of a cell, which is different from the survival of an organism, and his view of the genes of the simulated population as intelligent, though living genes are not (not even when they survive as producers of intelligent beings). When is an idea a mere metaphor and when does it become a decent theory?]

9. David L. Waltz, computer scientist, "The Prospects of Building Truly Intelligent Machines."

The old-style algoristically programmed AI machines incorporates a psychology that is now pass: the idea that all learning is by trial and error (p.195). [This regrettably is an exaggeration.] The crucial question is (p.196), "Given the immense range of possible situations a truly intelligent system could find itself in and the vast number of possible actions, how could the system ever manage to search out appropriate goals and actions?" [A false impression is given that the new style offers a solution to this question. Editor!] New-style parallel distributed machines learn to modify programs leading to any desired output [thus leaving the crucial question unanswered: how are goals generated?], have associative recall, and tolerate faults. The old-style machines do all this much more slowly than the new-style ones (p.199). The old-style algoristically programmed AI machines are taught each item unambiguously, making degrees of a computer's knowledge strictly a matter of quantities. This is logical [!], but infants have no explicit logic (p.201). A hybrid of old logical reasoning and new associative-memory learning machines are more human-like (pp.201 202).[Notice the false implicit equation of algoristic and logical thinking. Editor!]

The new-style machines work in "a process much more like lookup than search," in a process of looking up "items ... more like representations of specific or I episodes and objects than like rules and facts" (p.197). For example, a new-style "associative memory" diagnostic program prescribes lookups of records of previously diagnosed patients to select ones with symptoms most similar to those of the patient now under examination (p.198). Similarity or nearness between patterns is a statistical function averaging over a measure of distance defined between every pair of distinct items. The statistics may iron out faults, yet at the cost of uncertainty [especially since real lists of symptoms are usually much too short], so that the result of such a system is better checked by old methods (p.200). [In simulations; in live cases experts must check the computer's results. All this is irrelevant to the promised revolutionary learning psychology; the author even endorses Minsky's defunct psychologism; p. 201. How very very disappointing!] "Researchers have identified perhaps a dozen distinctly different learning methods" ; [this is very exciting, except that the author says nothing at all about any of them except to name] one of these [the psychologically least interesting] has an input and an output given in advance (p.204) [together with the initial program to be modified by trial and error!].

"The central problem ... in the connectionist" system is the "credit assignment problem": how should rewards and punishment to individual neuronlike elements be apportioned? [Relative to goals, of course. Goals were supposed to be created by the system itself, leaving no room for this problem. The central problem of the old system is thus reappearing in the new.] The "static part" of the problem is manageable [old-style]: units are tested upon their activities; when they act correctly their connections are strengthened and vice versa (p.205). The "temporal part" of the problem is more difficult as the system must remember its past states, to analyze and judge them, and then apportion rewards and punishments. [These are metaphors for the strengthening and weakening of the connections of the individual neuronlike items upon success or failure to perform adequately by set criteria. Taken literally, the metaphors are hilarious. Memory is here presented in the defunct associationist manner! The rest of the essay, perhaps most of it, is left out here; it defies summary; my main complaint, however, is that the author has a diffuse presentation of elementary learning psychology. Is this not a matter for the editor to have attended to?]

10. Anya Hurlbert, MD, AI researcher, and Tomasso Poggio, brain and cognitive scientist and researcher in computational vision, both in MIT, "Making Machines (and Artificial Intelligence) See."

"Why .. have we balked at calling vision intelligence?" If AI is to be interactive, it must include robotics, "the study of how to join perception with action ... In its beginning AI research ... excluded both vision and motor control from the realm of intelligence" (p.214).

Old-style AI follows the Newell-Simon [behaviorist and associationist] hypothesis that intelligence is "a physical symbol system," namely any computer. Being intelligent, then, humans are computers (p.215). By contrast,

new-style AI is Gestaltist. (p.214; cp. p.43, last note, on the influence of J.J. Gibson). "Leaps of intuition and instant insights are at one extreme, ordinary perceptual skills such as speech recognition at the other: these are the powers of the mind that traditional AI is hard put to model. ... Evolution has spent millennia perfecting such unconscious talents" (p.217). Machine vision is a synthesis of the best in AI, new style and old: "the computational approach"; it "describes exactly what information a system receives and what information it puts out, and seeks a computation that will transform the input into the output" within the recognized constraints of the normal living visual system. "Machine vision has turned the search for constraints into a science of the natural world"(p.218).

Most of the essay is devoted to the competent and sharp summary of the Marr-Poggio theory of vision (pp.218-230). [Whether it is too brief for the uninitiate or sufficiently detailed and succinct, it certainly cannot be further abbreviated here. Rather, what is needed is a clearer expression of what is so specific to the theory: the central role it ascribes to approximate solutions to problems. Here there is a confusion that can be found even in the writings of Sir Karl Popper: trial and error is not the same as approximationism: when a normal trial-and-error system is perfected, the details and the number of trials leading to the successful solution is insignificant; it may be beneficially reduced as long as (more often than not) improvement results. There are many trial-and-error algorisms that simply program to vary a solution and compare the new result with its predecessor and opt for the better one. More sophisticated trial-and error programs may exclude repetitions of older trials and at most have old outcomes rechecked. Approximationism is a trial-and-error system which does not discard earlier trials. The paradigm is still Einstein's use of Newtonian mechanics as an indispensable approximation. So is Newton's use of Kepler's and Galileo's results.]

[One of the most impressive experiments in vision theory is the Bela Julesz refutation of Gibson's theory of stereo-vision as the comparison of binocularly seen contours: in the [static] random stereogram depth is distinguishable in the stereo-vision of randomly distributed dots. To that end the observer first deems dots on the left and the right image close enough to be judged identical and only then their small local variations are distinguished and read as depth. Here the two steps are inherent in the process of intelligent surmise. In approximationism, unlike in most trial and error, the stages that are superseded are not discarded: recognizing them as partial achievements is part and parcel of the final recognition. This is nowadays recognized as a general aspect of pattern-recognition theory, which is not always agreeable: different paths may take an observer to different intelligent surmises of the same input, which is not acceptable. Yet this aspect of the situation is not specific to machines or to vision; it happens in diagnosis. This explains the proposal, made at times, to repeat a diagnostic process de novo. Farewell to associationism. The Marr-Poggio theory is systematically built with this idea in mind, yet with no accent put on the contrast between associations, trial and error in general, and approximations- in the Marr-Poggio paper and even in Marr's detailed book.] The penultimate section of the essay contrasts the three approaches, the algoristic old-style, the connectionist new-style and the computational revolutionary style. The third is that of the "true believer in levels of understanding." [Levels are approximations! See p.224.] The "associationist powers" are recognized by the connectionists; the "deductive powers" are recognized by the traditional computers; the computational view is thus a synthesis. [I do not understand this.]

The heart of the matter is "the single question: What is the final goal of the enterprise? ... is the goal ... to build intelligent machines? to understand how the brain is put together? to describe the structure and powers of intelligence as a free-floating entity, tied to neither brain nor machine?"(P.232.) The connectionist wants to have a model of the brain, seeing in the resemblance between humans and machines something abstract (p.233). Traditional AI, interested mainly in the construction of machines, will use results from brain physiology for convenience only. But its adherents, professing to be on the computational level, stay on the lower, algorismic level [where any algorism that does the job will suffice, in disregard for the total picture]. The connectionists only care to create the machine, ignoring all computations and forgetting that "many of the networks work only because the necessary computational analysis has been done first" (p.233).

An interesting example is taken up in the final passages of the essay: the behavior of a fly and the behavior of a driver in stress share the situation that their goals are not fixed. [This is an exaggeration: short-term goals depend on longer-term ones and on the (rapidly) altering conditions.] "Finding the right representation is what computational theory does. ... Machine vision shares the dream of building a machine that can learn, ... Will we be satisfied with simply building machines that can learn? ... We humans should not forget that those who aim to build intelligent machines have the whole future to disprove their starting hypothesis: that intelligence can be reproduced on a machine." [This is a serious error: the old AI theory is programmatic and thus metaphysical: it would be confirmed by success but not be refuted by failure, since there is "the whole future" to try yet again. It is a pity that these sophisticated researchers are not familiar with Popper's seminal methodological writings.] "..in the future more sophisticated machines ... might look fondly back at the days when machine vision, which combines all levels of understanding human intelligence, brought their parents together." 11. Sherry Turkle, MIT sociologist, "Artificial Intelligence and Psychoanalysis: A New Alliance."

"If psychoanalysis is in trouble, artificial intelligence may be able to help [and if psychoanalysis is not in trouble, AI will do no harm?] ... one of the ways computers influence psychological thinking is ... that ... computers provide science of mind with a kind of theoretical legitimation that I call sustaining myth" (pp.241-242). [Need one go further?]

Computer science helped kill behaviorism (pp.242-4) as well as the view of humans possess autonomous selves: the AI chess player "is more like Freud than Skinner" (pp.244-6, see also p.261ff.). Old style AI theoretician Marvin Minsky followed the old-style psychoanalysis of sigmund Freud (p.246; see also pp. 259-60 and note 17). "The two AIs, rule-driven and emergent, logical and biological in their aesthetic, fuel very different fantasies of how to build mind from machine." Old-style "information processing put AI in a distant relationship to psychoanalysis" and new style connectionist emergentism parallels the move within psychoanalysis from Freud to Melanie Klein in taking seriously object-relations so-called. Thus, "when the stuff of AI is expanded to include ... active and interactive inner agents, there is a starting place for a new dialogue between the psychoanalytic and the computer culture" (p.248). [This is superfluously grievously circumlocutious to the point of being meretricious: assuming that psychoanalysis describes humans correctly and

that AI tries to clone them, then, naturally, ... But cloning was never intended and if psychoanalysis is in trouble then the move to Kleinianism is either a reasonable way out in no need of AI legitimization or else ...]

12 and 13. Hilary Putnam, philosopher, "Much Ado About Not Very Much" and Daniel C. Dennett, philosopher, commenting on Putnam, "When Philosophers Encounter Artificial Intelligence."

Putnam: AI research is parasitic on computer science and its advocates mislead the public; natural languages and science are not given to algorisms and for the time being that is all there is to it. Dennett: the troubles Putnam mentions are well-known; he polarizes positions to all-or-nothing to deduce nothing from not-all; AI shows that memory problems involve storage and retrieval [this is a howler]; AI people test theories; as a philosopher Putnam likes neither tests nor AI.

[Dennett's paper is apologetic, ad hominem, and scarcely representative of his own accomplishments; he evades the challenge. Putnam thus wins an easy victory. Yet his judgment is facile, allowing not even for a silver lining. AI was hubris and deception; of necessity it did fail. Yet it did lead to some interesting efforts- even as failures, especially some technologically useful ones: whether expert systems are AI or computer programs is a matter not of definition but of history: where did the aspirations come from? Editor!]

14. John McCarthy, doyen of AI, "Mathematical Logic in Artificial Intelligence." [Comments are postponed.]

"A machine on the lowest level uses no logical sentences. It merely executes the commands on its program" (p.299). "The next level of logic use involves computer programs that put sentences in machine memory to represent their beliefs but use rules other than ordinary logical inferences to reach conclusions. New sentences are often obtained from old ones by ad hoc programs" (p.300). This is extremely limited, yet already good enough for expert systems. "The third level uses first-order logic as well as logical deduction" (p.300). "Examples ... used commercially are "expert-system shells" ... computer programs that create generic expert systems" (p.301). The third level is not so practical for lack of programs to induce intended deductions, especially ones of the most commonsense types, when these are defined (p.298) as deductions of sufficiently many simple everyday corollaries from given statements. Some progress in this direction has been achieved, and there still remains a fourth level, wholly in the future (pp.301-2): knowledge readable by any computer for any purpose. "The present way of "teaching" computers programs amounts to education by brain surgery."

The difficulty is in that "fourth-level systems require extensions to mathematical logic" (p.302). Traditional logic is monotonic: adding to the premises of a valid inference leaves it valid. But "some important human commonsense reasoning is not monotonic" (p.303). Some people try to save logic by the addition of rules of probability implicit in the context, but these are very doubtful (p.303). Agreeing with Quine that first-order logic should suffice and that there is no need to refer to ideas as we can scarcely say of two individuals that they have the same idea, the author nevertheless sees a need for a special logic (pp.303-5), one which will permit "jumping to conclusions on the basis of insufficient evidence" (p.307). He pleads for "incrementalism, or modesty" (p.307) which will permit AI researchers the privilege of trial and error, which we all have anyway. The example for trial and error he gives (p.308), however, is associationist: a baby first says "mother" thinking it is a singular and then learns it is a universal. The essay--and with it the book--ends with the hope that AI will help evolve some sort of meta-epistemology akin to traditional meta mathematics.

I do not know how to respond to this essay. I find it magnificent, broad, enlightening and simple overview. Yet I was taken aback by its confusions on matters of basic logic. I cannot complain: the errors show how inept we, the philosophical community, really are. But one may expect of a person of McCarthy's stature to know the best in logic rather than to be so confused.

The claim that ordinary reasonable thinking is non-monotonic is false. The examples McCarthy takes are not of violation of logic but of statements taken in the context in which they are stated, the context statements taken as stated and agreed upon, and of statements of context easily and naturally altered upon correction. This is not illogical. On the contrary, logic began as the theory of dialectic, and dialectic is the art of making explicit context statements and criticising them.

Thus, "jumping to conclusions on insufficient evidence" (p.307) is the norm--even though most theories of knowledge and of probability come to circumvent this fact. The meta-epistemology of the kind McCarthy seeks is already here: it is the theory of conjectures and refutations, of progress by trial and error; Karl Popper has presented it; the philosophical establishment is desperately overlooking it.

Logicians, too, are averse to trial and error. AI researchers are understandably in accord with computer scientists in their quite understandable penchant for formalism and for constructivism; but AI needs natural deduction theories--preferably Popper-style. (See Bibliography in P.A. Schilpp, ed., The Philosophy of Karl Popper.)

Nor are cognitive psychologists in a better state. Science, 30 October 1987, Volume 238, presents a paper on teaching reasoning by four famous cognitive psychologists who show that most people, including samples of science students, reason with highly brittle algorisms. They present alternatives similar to the ones presented by McCarthy for programming students with modes of reasoning. They too do not speak of trial and error as a mode of reasoning.

Let us take the rules of logic seriously, let us agree that associationism and inductivism are false, let us admit as a legitimate mode of reasoning [and as evolutionary] McCarthy's "incrementalism or modesty" ; let there be a blanket permission to AI researchers to engage in jumping to conclusions and critically looking at the results, in conjecturing and testing, in trial and error. Let us then have a program that boldly emulates (the baby and) the AI researcher.

The incremental attitude will easily show that the fourth level of computers (all-purpose multi-lingual meta-epistemological ones) that McCarthy says is wholly in the future, does exist, partly, already now, since computers can spot programmers' mistakes and since they can easily learn to spot formal contradictions. And there is a great need for fourth-level machines. There are partial expert systems diagnostic software programs on the market, and they have to be coordinated and merged into a comprehensive computer-assisted diagnostic service; this can only be achieved by a partial success of the fourth level. The service should also include competing expert systems. It is clear that commonsense is therefore essential. It is clear combining formal logic with commonsense requires

a better understanding of natural deduction, It is clear that interchangeable possible contexts--para-texts--for a given text, and rules for altering contexts are required and are available, at least partly, in current expert systems, but not systematically at all. Sets of contexts may require a meta-text that would embed some general metaphysical assumptions (see my "The Nature of Scientific Problems and their Roots in Metaphysics" in my Science in Flux, 1975) and some general technologically significant blanket suppositions (see my Technology, 1985). A first step would be the programming of a meta program for a brittle program to attempt tentatively different programs--different possible contexts--for the mending of brittleness for a while.

The idea behind this project (on which see more in Diagnosis by Nathaniel Laor and myself, NYUP, forthcoming [publication was for years prevented by the competitive interested parties]) is simple: there is no way to construct a fully artificial intelligence yet, nor is it moral to use one, as responsibility for action is always human. But there is the possibility to do so in limited contexts: dead languages and dead art can be made algorismic, and researchers can use computers in theirs researches only because they do formalize some of their procedures in some brittle ways.

In the very early days of computers (in the early 'fifties) Yehoshua Bar-Hillel has argued that totally mechanized machine translation is impossible, yet all the same he tried to formalize natural languages--taking it as a project not given to full success but worth-while anyway. It is, of course, impossible to feed a computer different levels of computation without the extensive use of a meta-linguistic program. As the computational approach does this already anyway, it seems clear that there is a prejudice against loading the meta-language and against second order logic that spills over to computer science and to AI. But clearly, inexact and undeveloped and perhaps as objectionable or redundant as W.V: Quine and John McCarthy say it is (p. 306), how come so many individuals are willing to emulate the human cognitive process even when it supposedly goes against logic and yet decline the use of heirarchies of languages and second-order logic and alternative frames to play with? Even non scientists non-logicians are known to do that! Why not view the "expert-systems shells" as brittle second order logic?

AI as the Golem myth is neither promising nor interesting; but AI as interactive wet-dry (or C/Fe) systems can rise to great highs: the story has not yet begun and there is no reason to despair. The chief issue raised by the editorial and the opening essay of this collection remains: does the allocation of financial resources in research in any way a significant contribution towards or away from significant progress of that research? Is research better off or worse of when publicly financed? If yes, is the present allocation wise? The few contrtibutions to this volume that touch upon the matter of finance--explicitly or delicately--suggests that of course the influence of money is always to the good. Usually, as long as not all financing of research into the matter is from the public sector, and as long as there is no public control over the research interests of genuinely curious academics, the question may very well be of little significance. Things get sensitive only in the exceptional cases, when research is extremely expensive and the private sector cannot or will not finance it but the military will. Even then it is not always deadly. Proof: we are not dead yet and even AI is alive and progressing despite the fact that the idea that science and art can be replaced by algorisms is obviously preposterous. This is not to say that public or private money foolishly squandered on research is no impediment: the greedy do block the free flow of information, especially the information critical of their activities. The worst is that their associates do that for them in good faith. Example. Science magazine is constantly bragging about its openness. Yet when the essay on the algorism called BACON by Simon and his associates was published there there was a flood of letters of criticisms (mine included) and they were all rejected by the judicious editors. In many circles, including some respected philosophy of science circles, this publication and its having remained uncontested was taken to represent a semi official view and one that invites reconsiderations. This is not the case, nor was there an intent to mislead, much less to suppress. Nevertheless. This story should, of course, light a little red bulb somewhere. It did not. Watch it.

Using Cautious Heuristics to Bias Generalization and Guide Example Selection

Diana F. Gordon
Navy Center for Applied Research in AI
Naval Research Laboratory, Code 5510
4555 Overlook Avenue, SW
Washington, D.C. 20375-5000
and
Department of Computer Science
University of Maryland
College Park, MD 20742

1. INTRODUCTION

Bias plays a significant role in inductive inference. In the framework of inductive concept learning from examples, there is an unknown target concept to be learned and a set of instances classified as positive or negative examples of the target concept. If learning is incremental, hypotheses (usually expressed in terms of instance features) are formed and then modified to remain consistent with the growing set of known instances. Generalization and specialization are frequently used for making modifications. A hypothesis is consistent with the instances if it logically implies all known positive instances and no known negative instances. If learning is empirical and the concept language is rich, the number of hypotheses consistent with the instances may be quite large. Since the purpose of each hypothesis is to predict over future instances, a judicious choice of one hypothesis over others ("bias")[1] can improve these predictions, thereby enhancing system performance.

In this paper, we discuss the use of explicit biases in the form of heuristics which recommend when to apply generalization operators, such as *drop-feature* and *climb-generalization-tree*, to incrementally modify hypotheses to learn a concept. These heuristics are based on definitions of conditions, such as feature *irrelevance*, which are important for learning. Heuristics offer two advantages. First, they bias hypothesis selection prior to hypothesis generation, which is a less computationally expensive method than using criteria to evaluate hypotheses that have already been generated.

---

[1] Called "inductive bias" in [Mitchell80] or "preference criteria" in [Michalski83].