ALCHEMY AND ARTIFICIAL INTELLIGENCE

Hubert L. Dreyfus

December 1965

P-3244

ALCHEMY AND ARTIFICIAL INTELLIGENCE

Hubert L. Dreyfus*

The RAND Corporation, Santa Monica, California

> The difference between the mathematical mind (esprit de géométrie) and the perceptive mind (esprit de finesse): the reason that mathematicians are not perceptive is that they do not see what is before them, and that, accustomed to the exact and plain principles of mathematics, and not reasoning till they have well inspected and arranged their principles, they are lost in matters of perception where the principles do not allow for such arrangement . . . . These principles are so fine and so numerous that a very delicate and very clear sense is needed to perceive them, and to judge rightly and justly when they are perceived, without for the most part being able to demonstrate them in order as in mathematics; because the principles are not known to us in the same way, and because it would be an endless matter to undertake it. We must see the matter at once, at one glance, and not by a process of reasoning, at least to a certain degree . . . . Mathematicians wish to treat matters of perception mathematically, and make themselves ridiculous . . . the mind . . . does it tacitly, naturally, and without technical rules.
>
> Pascal, Pènsées

---

*Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff.

This paper is based on an informal talk presented at The RAND Corporation in August 1964.

INTRODUCTION

Research dedicated to the construction of intelligent artifacts has, from its inception, intrigued philosophers, but thus far their discussions have been remarkably out of touch with the work actually being done. Analytic philosophers, such as Putnam, Scriven, and Ziff, use the present interest in "mechanical brains" to recast the conceptual issues dividing behaviorists from Cartesians. They assume that robots will eventually be built whose behavior will be indistinguishable from that of humans, and ask under what conditions we would be justified in saying that such an artifact was thinking. On the other hand, moralists and theologians evoke certain highly sophisticated forms of behavior--moral choice, love, creative abstraction, etc.--which they claim are beyond the powers of any machine. Neither side defines what sort of machine it has in mind nor tries to show that a machine can or cannot exhibit the behavior in question. Both parties credulously assume that highly intelligent artifacts have already been developed.

If such artifacts have been or are about to be produced, their operation will depend on the only high-speed, all-purpose information processing device which now exists--the digital computer. Thus, the only question which can reasonably be discussed at present is not whether robots can fall in love, or whether if they did we would say they were conscious, but rather to what extent a digital computer can be programmed to exhibit the sort of simple intelligent behavior characteristic of children and sometimes animals, such as playing games, solving simple

problems, reading sentences, and recognizing patterns. Philosophers have failed to raise this modest question. Instead, they approach the subject in terms of man's highest capacities, presumably because they are under the impression, fostered by the press and some artificial intelligence researchers, that these simple feats have been or are about to be performed. To begin with, then, these claims must be examined.

It is fitting to begin with a statement made in 1957 by H. A. Simon, one of the originators of the field of artificial intelligence:

> It is not my aim to surprise or shock you--if indeed that were possible in an age of nuclear fission and prospective interplanetary travel. But the simplest way I can summarize is to say that there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until--in a visible future--the range of problems they can handle will be coextensive with the range to which the human mind has been applied.

The speaker makes the following predictions:

1) That within ten years a digital computer will be the world's chess champion, unless the rules bar it from competition.

2) That within ten years a digital computer will discover and prove an important new mathematical theorem.

3) That within ten years a digital computer will write music that will be accepted by critics as possessing considerable aesthetic value.

> 4) That within ten years most theories in psychology will take the form of computer programs, or of qualitative statements about the characteristics of computer programs [34:7,8].*

Let us hope that in November 1967, the tenth anniversary of this historic talk, workers in the field of artificial intelligence will meet to measure their vision against reality. Meanwhile, though it is too early to definitively test these claims, enough time has elapsed to allow a significant confrontation of these predictions with actual progress in the field.

Recent publications suggest that the first of Simon's forecasts has already been half-realized and that considerable progress has been made in fulfilling his second prediction. In a review of Feigenbaum and Feldman's anthology, Computers and Thought, W. R. Ashby (one of the leading authorities in the field) hailed the mathematical power of of the properly programmed computer: "Gelernter's theorem-proving program has discovered a new proof of the pons asinorum that demands no construction." This proof, Professor Ashby goes on to say, is one which "the greatest mathematicians of 2000 years have failed to notice . . . which would have evoked the highest praise had it occurred" [2:2].

---

*References are listed alphabetically in the Bibliography at the end of this Paper. They are also numbered in this alphabetical order. Citations in the text are given in a bracketed pair of numbers: the first is the number of the reference itself, the second is the page on which the citation appears.

The theorem sounds important and the naive reader cannot help sharing Ashby's enthusiasm. A little research, however, reveals that the pons asinorum, or ass's bridge, is the first theorem to be proved in Euclidian geometry, viz., that the opposite angles of an isosceles triangle are equal. Moreover, the proof requiring the construction of a perpendicular to the base of the triangle (still taught in high schools) was introduced as late as the 19th century, presumably as a pedagogical device. The first announcement of the "new" proof "discovered" by the machine is attributed to Pappus (300 A.D.) [37:284]. There is a striking disparity between Ashby's excitement and the antiquity and triviality of this proof. We are still a long way from "the important mathematical theorem" to be found by 1967.

The chess-playing story is more involved and might serve as a model for a study of the production of intellectual smog in this area. The story began in 1955 with Allen Newell's sober survey of the problems posed by the game of chess and suggestions as to how they might be met. He found that "these [suggested] mechanisms are so complicated that it is impossible to predict whether they will work" [18:89].

The next year (a year before Simon makes his predictions) brought startling success. A group at Los Alamos produced a program which played poor but legal chess on a reduced board. In a review of this work, Newell, J. C. Shaw, and H. A. Simon concluded: "With very little in the way of complexity, we have at least entered the arena of human play--we can beat a beginner" [22:48]. In 1957, the

year of the great prediction, the Bernstein program for the IBM 704 entered the arena, and played two "passable amateur games" [22:45].

The following year, Newell, Shaw, and Simon (NSS) presented an elaborate chess-playing program. As described in their classic paper, "Chess Playing and the Problem of Complexity," their program was "not yet fully debugged," so that one "cannot say very much about the behavior of the program" [22:60]. Still, it is clearly "good in the opening." This is the last detailed published report on the program. In the same year, however, NSS announced: "We have written a program that plays chess" [21:6] and Simon, on the basis of this success, revised his earlier prediction.

> In another place, we have predicted that within ten years a computer will discover and prove an important mathematical theorem, and compose music that is regarded as aesthetically significant. On the basis of our experience with the heuristics of logic and chess, we are willing to add the further prediction that only moderate extrapolation is required from the capacities of programs already in existence to achieve the additional problem-solving power needed for such simulation [21:78].

In fact, in its few recorded games, the NSS program played poor but legal chess, and in its last official bout (October 1960) was beaten in 35 moves by a ten-year old novice. Fact, however, had ceased to be relevant. Newell, Shaw, and Simon's claims concerning their still bugged program had launched the chess machine into the realm of scientific mythology. In 1959, Norbert Wiener, whose optimism was strengthened by the claim that the program

was "good in the opening," informed the N.Y.U. Institute of Philosophy that "chess-playing machines as of now will counter the moves of a master game with the moves recognized as right in the text books, up to some point in the middle game" [41:110]. In the same symposium, Michael Scriven moved from the ambiguous claim that "machines now play chess" to the positive assertion that "machines are already capable of a good game" [32:128].

While their program was losing its five or six poor games--and the myth they had engendered was holding its own against masters in the middle game--Newell, Shaw, and Simon kept silent. When they speak again, three years later, they do not report their difficulties and disappointment. Rather, as if to take up where the myth had left off, Simon published an article in Behavioral Science announcing a program which will play "highly creative" chess end games involving "combinations as difficult as any that have been recorded in chess history" [36:429]. That the program restricts these end games to dependence on continuing checks, so that the number of relevant moves is greatly reduced, is mentioned but not emphasized. On the contrary, it is misleadingly implied that similar simple heuristics would account for master play even in the middle game. Thus, the article gives the impression that the chess prediction is almost realized. With such progress, the chess championship may be claimed at any moment. Indeed, a Russian cyberneticist, upon hearing of Simon's ten-year estimate, called it "conservative" [1:405]. And Fred Gruenberger at RAND has suggested that a world champion is not enough--that we should aim for "a program

which plays better than any man could" [11:6]. This output of confusion makes one think of the French mythical beast which is supposed to secrete the fog necessary for its own respiration.

I propose first to clear the air by reviewing the present state of artificial intelligence. The field has many divisions and subdivisions, but the most important work can be classified into four areas: a) game playing, b) problem solving, c) language translation and learning, and d) pattern recognition.

Part I will simply report the progress and difficulties in each area. Part II will show the common source of these seemingly unconnected difficulties and clarify certain conceptual confusions which hide the gravity of the situation these difficulties reveal. Part III will consider certain essential limitations on the information which can be processed by digital computers. Then, by classifying intelligent behavior in the light of these limitations, Part III will indicate which areas of behavior are susceptible to simulation and which areas lie beyond the capacities of digital computer programs.

Part I

THE CURRENT STATE OF THE FIELD OF ARTIFICIAL INTELLIGENCE

The field of artificial intelligence exhibits a recurrent pattern: early, dramatic success followed by sudden unexpected difficulties. Let us explore this pattern in detail.

SIGNS OF STAGNATION

Game Playing

The first years produced very impressive work--perhaps the most impressive work in the whole field of artificial intelligence. By 1955 Samuel had a checker program which could play "a fairly interesting game" [29:73]. After several improvements, including a learning program, Samuel's program was able to beat a former Connecticut checkers champion. Samuel's program does not attempt to simulate human information processing nor use heuristic search techniques. A tree of moves is searched to a depth which depends on the final position, and then, on the basis of an evaluation of certain parameters, a move is chosen.

This method is less successful in chess where the number of possible moves and responses is so great, the problem of exponential growth so acute, that the search tree must be pruned at each stage. Still, chess programs attained early success with simple limited search. The Los Alamos program, using no heuristics, could play a legal game on a reduced board. A year later, the Bernstein program using search pruning heuristics did as well on a full eight-by-eight

board. Then came the program developed by Newell, Shaw, and Simon, followed by the optimistic claims and predictions.

No one noted the unexpected difficulties. The initial NSS chess program was poor and, in the last five years, remains unimproved. Burton Bloom at M.I.T. has made the latest attempt to write a chess program; like all the others, it plays a stupid game. In fact, in the nine years since the Los Alamos program beat a weak player, in spite of a great investment of time, energy, and ink, the only improvement seems to be that a machine now plays poorly on an eight-by-eight rather than a six-by-six board. According to Newell, Shaw, and Simon themselves, evaluating the Los Alamos, the IBM, and the NSS programs: "All three programs play roughly the same quality of chess (mediocre) with roughly the same amount of computing time" [20:14]. Still no chess program can play even amateur chess, and the world championship tournament is only two years away.

Problem Solving

Again an early success: In 1957 Newell, Shaw, and Simon's Logic Theorist, using heuristically guided trial-and-error, proved 38 out of 52 theorems from Principia Mathematica. (Significantly, the greatest achievement in the field of mechanical theorem-proving, Wang's theorem-proving program, which proved in less than five minutes all 52 theorems chosen by Newell, Shaw, and Simon, does not use heuristics.) Two years later, the General Problem Solver (GPS), using more sophisticated means-ends analysis, solved the "cannibal and missionary" problem and other problems of similar complexity [22:15].

In 1961, after comparing a machine trace with a protocol which matched the machine output to some extent, Newell and Simon concluded rather cautiously:

> The fragmentary evidence we have obtained to date encourages us to think that the General Problem Solver provides a rather good first approximation to an information processing theory of certain kinds of thinking and problem solving behavior. The processes of "thinking" can no longer be regarded as completely mysterious (my italics) [24:19].

Soon, however, Simon gave way to more enthusiastic claims:

> Subsequent work has tended to confirm [our] initial hunch, and to demonstrate that heuristics, or rules of thumb, form the integral core of human problem-solving processes. As we begin to understand the nature of the heuristics that people use in thinking, the mystery begins to dissolve from such (heretofore) vaguely understood processes as "intuition" and "judgment" [33:12].

But, as we have seen in the case of chess, difficulties have an annoying way of reasserting themselves. This time, the "mystery" of judgment reappears in terms of the organizational aspects of the problem-solving programs. In "Some Problems of Basic Organization in Problem-Solving Programs" (December 1962), Newell discusses some of the problems which arise in organizing the Chess Program, the Logic Theorist, and especially the GPS, with a candor rare in the field, and admits that "most of them are unsolved to some extent, either completely, or because the solutions that have been adopted are still unsatisfactory in one way or another" [19:4]. No further progress has been reported toward the resolution of these problems.

This curve from success to optimism to disappointment can be followed in miniature in the case of Gelernter's Geometry Theorem Machine (1959). Its early success with theorems like the pons asinorum gave rise to the first prediction sufficiently short-range to have already been totally discredited. In an article published in 1960, Gelernter explains the heuristics of his program and then concludes: "Three years ago, the dominant opinion was that the geometry machine would not exist today. And today, hardly an expert will contest the assertion that machines will be proving interesting theorems in number theory three years hence," i.e., in 1963 [9:160]. No more striking example exists of an "astonishing" early success and the equally astonishing failure to follow it up.

Language Translation

This area had the earliest success, the most extensive and expensive research, and the most unequivocal failure. It was clear from the start that a mechanical dictionary could easily be constructed in which linguistic items, whether they were parts of words, whole words, or groups of words, could be processed independently and converted one after another into corresponding items in another language. As Richard See notes in his article in Science, May 1964: "Successful processing at this most primitive level was achieved at an early date" [30:622], and Oettinger, the first to produce a mechanical dictionary (1954), recalls this early enthusiasm: "The notion of . . . fully automatic high quality mechanical translation, planted by over-zealous propagandists for automatic translation on both sides of the

Iron Curtain and nurtured by the wishful thinking of potential users, blossomed like a vigorous weed" [27:18]. This initial success and the subsequent disillusionment provides a sort of paradigm for the field. It is aptly described by Bar-Hillel in his report on "The Present Status of Automatic Translation of Languages."

> During the first year of the research in machine translation, a considerable amount of progress was made . . . . It created among many of the workers actively engaged in this field the strong feeling that a working system was just around the corner. Though it is understandable that such an illusion should have been formed at the time, it was an illusion. It was created . . . by the fact that a large number of problems were rather readily solved . . . . It was not sufficiently realized that the gap between such output . . . and high quality translation proper was still enormous, and that the problems solved until then were indeed many but just the simplest ones whereas the "few" remaining problems were the harder ones--very hard indeed [3:94].

During the ten years since the development of a mechanical dictionary, five government agencies have spent about 16 million dollars on mechanical translation research [30:625]. In spite of journalistic claims at various moments that machine translation was at last operational, this research produced primarily a much deeper knowledge of the unsuspected complexity of syntax and semantics. As Oettinger remarks, "The major problem of selecting an appropriate target correspondent for a source word on the basis of context remains unsolved, as does the related one of establishing a unique syntactic structure for a sentence that human readers find unambiguous" [27:21]. Oettinger

concludes: "The outlook is grim for those who still cherish hopes for fully automatic high-quality mechanical translation" [27:27]. Acting on Oettinger's realization, the Harvard Computation Laboratory decided to concentrate its work on English syntax and dropped all work on Russian.

Pattern Recognition

This field is discussed last because the resolution of the difficulties which have arrested development in game playing, problem solving, and language translation all presuppose success in the field of pattern recognition (which in turn suffers from each of the difficulties encountered in the other fields). As Selfridge and Neisser point out in their classic article, "Pattern Recognition by Machine,"

> . . . a man is continually exposed to a welter of data from his senses, and abstracts from it the patterns relevant to his activity at the moment. His ability to solve problems, prove theorems and generally run his life depends on this type of perception. We suspect that until programs to perceive patterns can be developed, achievements in mechanical problem-solving will remain isolated technical triumphs [31:238].

As one might expect, this field experienced no simple early successes. Selfridge and Neisser allow that "developing pattern recognition programs has proved rather difficult" [31:238]. There has indeed been some excellent work. The Lincoln Laboratory group under Bernard Gold produced a program for transliterating hand-sent Morse code. And there are several operational programs that can learn to recognize hand-printed alphabetic characters of

variable sizes and rotations. Still, as Selfridge and Neisser remark, "At present the only way the machine can get an adequate set of features is from a human programmer" [31:244]. And they conclude their survey of the field with a challenge rather than a prediction:

> The most important learning process of all is still untouched: No current program can generate test features of its own. The effectiveness of all of them is forever restricted by the ingenuity or arbitrariness of their programmers. We can barely guess how this restriction might be overcome. Until it is, "artificial intelligence" will remain tainted with artifice [31:250].

Even these remarks may be too optimistic, however, in their supposition that the present problem is feature-generation. The relative success of the Uhr-Vossler program, which generates and evaluates its own operators, shows that this problem is partially soluble. However, as Part II demonstrates, mechanical recognition still remains a rigid process of brute-force enumeration. No pattern recognition program, even Uhr-Vossler's, incorporates the flexibility of the human tacit pattern-recognition processes.

Thus the disparity between prediction and performance which is characteristic of artificial intelligence re-appears, for example, in the case of the print reader. Bar-Hillel, who also likes to collect unfulfilled prophesies, quoted in 1959 a claim by Edwin Reifler that "in about two years [from August 1957] we shall have a device which will at one glance read a whole page." Bar-Hillel, who presumably has been cured of over-optimism, then went on to make a more modest claim. "The best estimates I am aware of at

present mention five years as the time after which we are likely to have a reliable and versatile print reader . . ." [3:104].

Over five years have elapsed since Bar-Hillel made this conservative estimate. At that time, flight to the moon was still science fiction and the print reader was just around the corner. Now the moon project is well underway while, according to Oettinger in 1963, no versatile print reader is in sight: "In the foreseeable future, automatic print reading devices will handle only materials with great uniformity of layout and type design, such as, for example, ordinary typewritten material" [27:20]. Books and journals with registered margins seem to be over the horizon, and the horizon seems to be receding at an accelerating rate.

Comments and Conclusions

An overall pattern is taking shape: an early, dramatic success based on the easy performance of simple tasks, or low-quality work on complex tasks, and then diminishing returns, disenchantment, and, in some cases, pessimism. The pattern is not caused by too much being demanded too soon by eager or skeptical outsiders. The failure to produce is measured solely against the expectations of those working in the field.

When the situation is grim, however, enthusiasts can always fall back on their own optimism. This tendency to substitute long-range for operational programs slips out in Feigenbaum and Feldman's claim that "the forecast for progress in research in human cognitive processes is most

encouraging" [8:276]. The forecast always has been, but one wonders: how encouraging are the prospects?

Feigenbaum and Feldman claim that tangible progress is indeed being made and they define progress very carefully as "displacement toward the ultimate goal" [8:vi]. According to this definition, the first man to climb a tree could claim tangible progress toward flight to the moon.*

---

*An example of the absurdity to which this notion of progress leads is the suggestion that the baseball program which answers questions posed in a drastically restricted vocabulary and syntax is an "important initial step toward [the] goal . . . of discovering the information processing structure underlying the act of 'comprehending' or the process of 'understanding'" [8:205].

Part II

THE UNDERLYING SIGNIFICANCE OF CURRENT DIFFICULTIES

Negative results can be interesting, provided one recognizes them as such. The diminishing achievement, instead of the predicted accelerating success, perhaps indicates some unexpected phenomenon. Are we pushing out on a continuum like that of velocity, such that progress becomes more and more difficult as we approach the speed of light, or are we instead facing a discontinuity, like the tree-climbing man who reaches for the moon?

It seems natural to take stock of the field at this point, yet surprisingly no one has done so. If someone had, he would have found that each of the four areas considered has a corresponding specific form of human information processing which enables human subjects in that area to avoid the difficulties which an artificial subject must confront. The section below will isolate these four human forms of information processing and contrast them with their machine surrogates. The following section will formulate and criticize the assumption shared by workers in artificial intelligence that human subjects face the same difficulties as artificial subjects and that therefore these difficulties obviously can be overcome.

HUMAN VS. MACHINE INFORMATION PROCESSING

Fringe Consciousness Vs. Heuristically Guided Search

It is common knowledge that a certain class of games are decidable on present-day computers with present-day

techniques--games like nim and tic-tac-toe can be programmed so that the machine will win or draw every time. Other games, however, cannot be decided in this way on present-day computers, and yet have been successfully programmed. In checkers, for example, because only two kinds of moves are possible, the captures are forced, and pieces block each other, one can explore all possibilities to a depth of as many as twenty moves, which proves sufficient for playing a good game.

Chess, however, presents the problem inevitably connected with choice mazes: exponential growth. We cannot run through all the branching possibilities even far enough to form a reliable judgment as to whether a given branch is sufficiently promising to merit further exploration. Newell notes that it would take much too long to find an interesting move if the machine had to examine the pieces on the board one after another. He is also aware that, if this is not done, the machine may sometimes miss an important and original combination. "We do not want the machine to spend all its time examining the future actions of committed men; yet if it were never to do this, it could overlook real opportunities . . ." [18:80].

His first solution was "the random element . . .. The machine should rarely [i.e., occasionally] search for combinations which sacrifice a Queen . . ." [18:80]. But this solution is unsatisfactory, as Newell himself presumably now realizes. The machine should not look just every once in a while for a Queen sacrifice but, rather, look in those situations in which such a sacrifice would be relevant. This is what the right heuristics are supposed

to assure, by limiting the number of branches explored while retaining the more promising alternatives.

No such heuristics have as yet been found. All current heuristics either exclude some possibly good moves or leave open the risk of exponential growth. Simon is nonetheless convinced, for reasons discussed below, that chess masters use such heuristics, and so he is confident that if we listen to their protocols, follow their eye movements, perhaps question them under bright lights, we can eventually discover these heuristics and build them into our program--thereby pruning the exponential tree. But let us examine more closely the evidence that chess playing is governed by the use of heuristics.

Consider the following protocol quoted by Simon, noting especially how it begins rather than how it ends. The player says,

> Again I notice that one of his pieces is not defended, the Rook, and there must be ways of taking advantage of this. Suppose now, if I push the pawn up at Bishop four, if the Bishop retreats I have a Queen check and I can pick up the Rook. If, etc., etc. [24:15].

At the end we have an example of what I shall call "counting out"--thinking through the various possibilities by brute-force enumeration. We have all engaged in this process, which, guided by suitable heuristics, is supposed to account for the performance of chess masters. But how did our subject notice that the opponent's Rook was undefended? Did he examine each of his opponent's pieces and their possible defenders sequentially (or simultaneously) until he stumbled on the vulnerable Rook? Impossible! As

Newell, Shaw, and Simon remark, "The best evidence suggests that a human player considers considerably less than 100 positions in the analysis of a move" [22:47], and our player must still consider many positions in evaluating the situation once the undefended Rook has been discovered.

We need not appeal to introspection to discover what a player in fact does before he begins to count out; the protocol itself indicates it: the subject "zeroed in" on the promising situation ("I notice that one of his pieces is not defended"). Often, of course, locating the promising or threatening area involves more than simply noticing that a Rook is undefended. It may involve noticing that "here something interesting seems to be going on"; "he looks weak over here"; "I look weak over there"; etc. Only after the player has zeroed in on an area does he begin to count out, to test, what he can do from there.

The player need not be aware of having explicitly considered or explicitly excluded from consideration any of the hundreds of possibilities that would have had to be enumerated in order to have arrived at this particular area by counting out. Still, the specific portion of the board which finally attracts the subject's attention depends on the overall configuration. To understand how this is possible, consider what William James has called "the fringes of consciousness": the ticking of a clock which we notice only if it stops provides a simple example of this sort of marginal awareness. Our vague awareness of the faces in a crowd when we search for a friend is another, more complex and more nearly appropriate, case.

But in neither of these cases does the subject make positive use of the information resting on the fringe. The chess case is best understood in terms of Polanyi's description of the power of the fringes of consciousness to concentrate information concerning our peripheral experience.

> This power resides in the area which tends to function as a background because it extends indeterminately around the central object of our attention. Seen thus from the corner of our eyes, or remembered at the back of our mind, this area compellingly affects the way we see the object on which we are focusing. We may indeed go so far as to say that we are aware of this subsidiarily noticed area mainly in the appearance of the object to which we are attending [28:214].

Once familiar with a house, for example, the front looks thicker than a facade, because one is marginally aware of the house behind. Similarly, in chess, cues from all over the board, while remaining on the fringes of consciousness, draw attention to certain sectors by making them appear promising, dangerous, or simply worth looking into.

If information, rather than being explicitly considered, can remain on the fringes of consciousness and be implicitly taken into account through its effect on the appearance of the objects on which our attention is focused, then there is no reason to suppose that, in order to discover an undefended Rook, our subject must have counted out rapidly and unconsciously until he arrived at the area in which he began consciously counting out. Moreover, there are good reasons to reject this assumption, since it raises more problems than it solves.

If the subject has been unconsciously counting out thousands of alternatives with brilliant heuristics to get to the point where he focuses on that Rook, why doesn't he carry on with that unconscious process all the way to the end, until the best move just pops into his consciousness? Why, if the unconscious counting is rapid and accurate, does he resort at the particular point where he spots the Rook to a cumbersome method of slowly, awkwardly, and consciously counting things out? Or if, on the other hand, the unconscious counting is inadequate, what is the advantage of switching to a conscious version of the same process?

It seems that "unconsciously" the subject is engaged in a sort of information processing which differs from counting out, and conscious counting begins when he has to refine this global process in order to deal with details. Moreover, even if he does unconsciously count out, using unconscious heuristics--which there is no reason to suppose and good reason to doubt--what kind of program could convert this unconscious counting into the kind of fring-influenced awareness of the centers of interest, which is the way zeroing-in presents itself in our experience? Why has no one interested in cognitive simulation been interested in this conversion process?

There is thus no evidence, behavioral or introspective, that counting out is the only function of thought involved in playing chess, that "the essential nature of the task [is] search in a space of exponentially growing possibilities" [22:65]. On the contrary, all protocols testify that chess involves two kinds of behavior: zeroing in on

an area formerly on the fringes of consciousness, which other areas still on the fringes of consciousness make interesting; and counting out explicit alternatives.

This distinction clarifies the early success and the later failure of work in artificial intelligence. In all game-playing programs, early success is attained by working on those games or parts of games in which counting out is feasible; failure occurs when global awareness is necessary to avoid exponential growth.

Essence/Accident Discrimination vs. Trial and Error

Work in problem solving also encounters two functions of thought--one, elementary and associationistic, accounts for the early success in the field; another, more complex and requiring insight, has proved intractable to step-wise programs such as the GPS.

If a problem is set up in a simple, completely determinate way, with an end and a beginning and rules for getting from one to the other (in other words, if we have what Simon calls a "simple formal problem"), then GPS can successfully bring the end and the beginning closer and closer together until the problem is solved. But even this presents many difficulties. Comparing the trace of a GPS solution with the protocol of a human solving the same problem reveals steps in the machine trace (explicit searching) which do not appear in the subject's protocol. And we are again asked to accept the dubious assumption that "many things concerning the task surely occurred without the subject's commenting on them (or being aware of them)" [26:288], and the even more arbitrary

assumption that these further operations were of the same elementary sort as those verbalized. In fact, certain details of Newell and Simon's article, "GPS: A Program that Simulates Human Thought," suggest that these further operations are not like the programmed operations at all.

At a point in the protocol analyzed in this article, the subject applies the rule $(A \cdot B \rightarrow A,\ A \cdot B \rightarrow B)$, to the conjunction $(\sim R \vee \sim P) \cdot (R \vee Q)$. Newell and Simon note:

> The subject handled both forms of rule 8 together, at least as far as his comment is concerned. GPS, on the other hand, took a separate cycle of consideration for each form. Possibly the subject followed the program covertly and simply reported the two results together [26:289].

Probably, however, the subject grasped the conjunction as symmetric with respect to the transformation operated by the rule, and so in fact applied both forms of the rule at once. Even Newell and Simon admit that they would have preferred that GPS apply both forms of the rule in the same cycle. They wisely refrain, however, from trying to write a program which could discriminate between occasions when it was appropriate to apply both forms of the rule at once and those when it was not. Such a program, far from eliminating the above divergence, would require further processing not reported by the subject, thereby increasing the discrepancy between the program and the protocol. Unable thus to eliminate the divergence and unwilling to try to understand its significance, Newell and Simon dispose of the discrepancy as "an example of parallel processing" [26:290].

Another divergence noted by Newell and Simon, however, does not permit such an evasion. At a certain point, the protocol reads: ". . . I should have used rule 6 on the left-hand side of the equation. So use 6, but only on the left-hand side." Simon notes:

> Here we have a strong departure from the GPS trace. Both the subject and GPS found rule 6 as the appropriate one to change signs. At this point GPS simply applied the rule to the current expression; whereas the subject went back and corrected the previous application. Nothing exists in the program that corresponds to this. The most direct explanation is that the application of rule 6 in the inverse direction is perceived by the subject as undoing the previous application of rule 6 [26:291].

This is indeed the most direct explanation, but Newell and Simon do not seem to realize that this departure from the trace, which cannot be explained away by parallel processing, is as serious as the planetary discrepancies which alerted modern astronomers to the inadequacies of the Ptolemaic system. Some form of thinking other than searching is taking place.

Newell and Simon note the problem: "It clearly implies a mechanism [maybe a whole set of them] that is not in GPS" [26:292], but, like the ancient astronomers, they try to save their theory by adding a few epicycles. They continue to suppose, without any evidence, that this mechanism is just a more elaborate search technique which can be accommodated by providing GPS with "a little continuous hindsight about its past actions" [26:292]. They do not realize that their subject's decision to backtrack must be the result of a very <u>selective</u> checking procedure.

Otherwise, all past steps would have to be rechecked at each stage, which would hopelessly encumber the program.

A more scientific approach would be to explore further the implications of the five discrepancies noted in the article, in order to determine whether or not a different form of information processing might be involved. For example, Wertheimer points out in his classic work, Productive Thinking, that the associationist account of problem solving excludes the most important aspect of problem solving behavior, viz., a grasp of the essential structure of the problem, which he calls "insight" [40:202]. In this operation, one breaks away from the surface structure and sees the basic problem--what Wertheimer calls the "deeper structure"--which enables one to organize the steps necessary for a solution.

This gestaltist conception may seem antithetical to the operational concepts demanded in artificial intelligence, but in fact this restructuring is surreptitiously presupposed by the work of Newell, Shaw, and Simon themselves. In The Processes of Creative Thinking, they introduce "the heuristics of planning" to account for characteristics of the subject's protocol lacking in a simple means-end analysis.

> We have devised a program . . . to describe the way some of our subjects handle O. K. Moore's logic problems, and perhaps the easiest way to show what is involved in planning is to describe that program. On a purely pragmatic basis, the twelve operators that are admitted in this system of logic can be put in two classes, which we shall call "essential" and "inessential" operators, respectively. Essential operators are those which, when applied to an expression, make "large"

> changes in its appearance--change "PvP" to "P", for example. Inessential operators are those which make "small" changes--e.g., change "PvQ" to "QvP". As we have said, the distinction is purely pragmatic. Of the twelve operators in this calculus, we have classified eight as essential and four as inessential . . . .
>
> Next, we can take an expression and abstract from it those features that relate only to essential changes. For example, we can abstract from "PvQ" the expression (PQ), where the order of the symbols in the latter expression is regarded as irrelevant. Clearly, if inessential operations are applied to the abstracted expressions, the expressions will remain unchanged, while essential operations can be expected to change them . . . .
>
> We can now set up a correspondence between our original expressions and operators, on the one hand, and the abstracted expressions and essential operators, on the other. Corresponding to the original problem of transforming a into b, we can construct a new problem of transforming a' into b', where a' and b' are the expressions obtained by abstracting a and b respectively. Suppose that we solve the new problem, obtaining a sequence of expressions, a'c'd' ...b'. We can now transform back to the original problem space and set up the new problems of transforming a into c, c into d, and so on. Thus, the solution of the problem in the planning space provides a plan for the solution of the original problem [21:43,44].

No comment is necessary. One merely has to note that the actual program description begins in the second paragraph. The classification of the operators into essential and inessential, the function Wertheimer calls "finding the deeper structure" or "insight," is introduced by the programmers before the actual programming begins.

This human ability to distinguish the accidental from the essential accounts for the divergence of the protocol of the problem-solving subjects from the machine trace. We have already suggested that the subject applies both forms of rule 8 together because he realizes that, at this initial stage, both sides of the conjunction are functionally equivalent. Likewise, because he has grasped the essential function of rule 6, the subject can see that the present application of the rule simply neutralizes the previous one. As Wertheimer notes:

> The process [of structuring a problem] does not involve merely the given parts and their transformations. It works in conjunction with material that is structurally relevant but is selected from past experience . . . [40:195].

No one has even tried to suggest how a machine could perform this structuring operation or how it could be learned, since it is one of the conditions for learning from past experience. The ability to distinguish the essential from the inessential seems to be a uniquely human form of information processing not amenable to the mechanical search techniques, which may operate once this distinction has been made. It is precisely this function of intelligence which resists further progress in the problem-solving field.

In the light of their frank recourse to the insightful predigesting of their material, there seems to be no foundation for Newell, Shaw, and Simon's claim that the behavior vaguely labeled cleverness or keen insight in human problem solving is really just the result of the judicious application of certain heuristics for narrowing

and guiding the search for solutions. Their work on GPS, on the contrary, demonstrates that all searching, unless directed by a preliminary structuring of the problem, is merely a blind muddling through.

Ironically, research in cognitive simulation is the only example of so-called intelligent behavior which proceeds like the unaided GPS. Here one finds the kind of muddling through and ad hoc patching up characteristic of a fascination with the surface structure--a sort of tree-climbing with one's eyes on the moon. Perhaps because the field provides no example of insight, some people in cognitive simulation have mistaken the operation of GPS for intelligent behavior.

Ambiguity Tolerance vs. Exhaustive Enumeration

Work on game playing revealed the necessity of processing information which is not explicitly considered or rejected, i.e., information on the fringes of consciousness. Problem solving research demonstrated that a distinction between the essential and the accidental is presupposed in attacking a problem. Work in language translation has been halted by the need for a third, non-programmable form of information processing.

We have seen that Bar-Hillel and Oettinger, two of the most respected and best informed workers in the field of automatic language translation, have been led to similar pessimistic conclusions concerning the possibility of further progress in the field. They have each realized that, in order to translate a natural language, more is needed than a mechanical dictionary, no matter how complete,

and the laws of grammar, no matter how sophisticated. The order of the words in a sentence does not provide enough information to enable a machine to determine which of several possible parsings is the appropriate one, nor does the context of a word indicate which of several possible meanings is the one the author had in mind.

As Oettinger says in discussing systems for producing all parsings of a sentence acceptable to a given grammar:

> The operation of such analyzors to date has revealed a far higher degree of legitimate syntactic ambiguity in English and in Russian than has been anticipated. This, and a related fuzziness of the boundary between the grammatical and the non-grammatical, raises serious questions about the possibility of effective fully automatic manipulation of English or Russian for any purposes of translation or information retrieval [27:26].

Instead of claiming, on the basis of his early partial success with a mechanical dictionary, that, in spite of a few exceptions and difficulties, the mystery surrounding our understanding of language is beginning to dissolve, Oettinger draws attention to the "very mysterious semantic processes that enable most reasonable people to interpret most reasonable sentences unequivocally most of the time . . ." [27:26].

Here is another example of the importance of the fringe effect. Obviously, the user of a natural language is not aware of many of the cues to which he responds in determining the intended syntax and meaning. On the other hand, nothing indicates that he considers each of these cues unconsciously. In fact, two considerations suggest that these cues are not the sort that could be taken up and

considered by a sequential or even parallel list-searching program.

First, too many possibly relevant cues exist, as Bar-Hillel concludes in an argument "which amounts to an almost full-fledged demonstration of the unattainability of fully automatic high quality translation, not only in the near future but altogether" [3:94]. The argument is sufficiently important to merit quoting at some length.

> I shall show that there exist extremely simple sentences in English--and the same holds, I am sure, for any other natural language--which, within certain linguistic contexts, would be uniquely (up to plain synonymy) and unambiguously translated into any other language by anyone with a sufficient knowledge of the two languages involved, though I know of no program that would enable a machine to come up with this unique rendering unless by a completely arbitrary and ad hoc procedure whose futility would show itself in the next example.
>
> A sentence of this kind is the following:
>
> The box was in the pen.
>
> The linguistic context from which this sentence is taken is, say, the following:
>
> > Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy.
>
> Assume, for simplicity's sake, that pen in English has only the following two meanings: (1) a certain writing utensil, (2) an enclosure where small children can play. I now claim that no existing or imaginable program will enable an electronic computer to determine that the word pen in the given sentence within the given context has the second of the above meanings, whereas every reader with a sufficient knowledge of English will do this "automatically" [3:158,159].
>
> What makes an intelligent human reader grasp this meaning so unhesitatingly is, in addition to all the other features that have been discussed

> by MT workers . . . , his knowledge that the relative sizes of pens, in the sense of writing implements, toy boxes, and pens, in the sense of playpens, are such that when someone writes under ordinary circumstances and in something like the given context, "The box was in the pen," he almost certainly refers to a playpen and most certainly not to a writing pen [3:160].

And, as Bar-Hillel goes on to argue, the suggestion that a computer used in translating be supplied with a universal encyclopedia is "utterly chimerical." "The number of facts we human beings know is, in a certain very pregnant sense, infinite" [3:160]. Even if the number of facts was only very large and even if all these facts could be stored in an enormous list in our memory or in a machine, neither we nor the machine could possibly search such a list in order to resolve semantic and syntactic ambiguities.

Second, even if a manageable number of relevant cues existed, they would not help us: in order to use a computer to interpret these cues, we would have to formulate syntactic and semantic criteria in terms of strict rules; and our use of language, while precise, is not strictly rule-like. Pascal already noted that the perceptive mind functions "tacitly, naturally, and without technical rules." Wittgenstein has spelled out this insight in the case of language.

> We are unable clearly to circumscribe the concepts we use; not because we don't know their real definition, but because there is no real "definition" to them. To suppose that there must be would be like supposing that whenever children

> play with a ball they play a game according to strict rules [43:25].*

A natural language is used by people involved in situations in which they are pursuing certain goals. These extra-linguistic goals, which need not themselves be precisely stated or statable, provide the cues which reduce the ambiguity of expressions as much as is necessary for the task at hand. A phrase like "stand near me" can mean anything from "press up against me" to "stand one mile away," depending upon whether it is addressed to a child in a crowd or to a fellow scientist at Los Alamos. Even in context its meaning is imprecise, but it is precise enough to get the intended result.

Our ability to use a global context to <u>sufficiently reduce ambiguity</u> without having to formalize (i.e., eliminate ambiguity altogether), reveals a third fundamental form of human information processing, which presupposes the other two. Fringe consciousness makes us aware of cues in the context which are too numerous to be made explicit. A pragmatic sense of what is essential in a given context allows us to ignore as irrelevant certain possible parsings of sentences and meanings of words which would be included in the output of a machine. Ambiguity

---

*The participants in the RAND symposium on "Computers and Comprehension" suggest the psychological basis and advantage of this non-rule-like character of natural languages.

> It is crucial that language is a combinatory repertoire with unlimited possible combinations whose meanings can be inferred from a finite set of "rules" governing the components' meaning. (The so-called "rules" are learned as response sets and are only partly formalizable.) [13:12]

tolerance then allows us to use this information about goals and context to narrow down the remaining spectrum of possible parsings and meanings as much as the situation requires without requiring the resulting interpretation to be absolutely unambiguous.

Since understanding a sentence in a natural language requires a knowledge of extra-linguistic facts and a grasp of the sentence's context-dependent use--neither of which we learn from explicit rules--the only way to make a computer which could understand and translate a natural language is to program it to learn about the world. Bar-Hillel remarks: "I do not believe that machines whose programs do not enable them to learn, in a sophisticated sense of this word, will ever be able to consistently produce high-quality translations" [3:105,106].*

In the area of language-learning, the only interesting and successful program is Feigenbaum's EPAM (Elementary Perceiver and Memorizer). EPAM simulates the learning of the association of nonsense syllables, which Feigenbaum calls "a simplified case of language learning" [7:289].

The interesting thing about nonsense syllable learning, however, is that it is not a case of language learning at all. Learning to associate nonsense syllables is in fact acquiring a Pavlovian conditioned reflex. The machine could exhibit "DAX" then "JIR" or it could flash red and then green lights; as long as two such events were

---

*Among workers in artificial intelligence, only MacKay has made specific suggestions as to what form such "sophisticated learning" programs might take (cf., "An Internal Representation of the External World" [15]).

associated frequently enough, one would learn to anticipate the second member of the pair. In such an experiment, the subject is supposed to be completely passive. In a sense, he isn't really learning anything, but is having something done to him. Whether the subject is an idiot, a child, or a genius should ideally make no difference in the case of nonsense syllable learning. Ebenhouse, at the end of the 19th century, proposed this form of conditioning precisely to eliminate any use of meaningful grouping or appeal to a context of previously learned associations.

It is no surprise that subject protocol and machine trace most nearly match in this area. But it is a dubious triumph: the only successful case of cognitive simulation simulates a process which does not involve comprehension and so is not genuinely cognitive.

What is involved in learning a language is much more complicated, and more mysterious, than the sort of conditioned reflex involved in learning to associate nonsense syllables. To teach someone the meaning of a new word, we can sometimes point at the object which the word names. Since Augustine's Confessions, it has been assumed that this is the way we teach language to children. But Wittgenstein pointed out that if we simply point at a table, for example, and say "brown," a child may not known if brown is the color, the size or the shape of the table, the kind of object, or the proper name of the object. If the child already uses language, we can say that we are pointing out the color, but if he doesn't already use language, how do we ever get off the ground? Wittgenstein says that the subject must be engaged in a form of life in

which he shares at least some of the goals and interests of the teacher, so that the activity at hand helps determine the meanings of the words used.

The above considerations concerning the essential role of context awareness and ambiguity tolerance in the use of a natural language should suggest why work is coming to a halt in the translating field. Furthermore, the ability to learn a language presupposes a complex combination of the uniquely human forms of information processing, so that an appeal to learning cannot be used to bypass the problems confronting this area.

Perspicuous Grouping--A Derivative of the Above Three Forms

Successful recognition of even simple patterns requires each of the fundamental forms of human information processing discussed thus far; recognition of patterns as complex as artistic styles and the human face requires, in addition, a special combination of the above three. It is no wonder that work in pattern recognition has had a late start and an early stagnation.

Part I noted that a weakness of current pattern recognition programs (with the possible exception of the Uhr-Vossler program, the power of whose operators--since it only recognizes five letters--has not yet been sufficiently tested) is that they are not able to determine their own selection operators. Now, however, we shall see that this way of presenting the problem is based on assumptions which hide deeper and more difficult issues.

Insight. A first indication that human pattern recognition differs radically from mechanical recognition is

seen in human (and animal) tolerance for changes in orientation and size, degrees of incompleteness and distortion, and amount of background noise.

An early artificial intelligence approach was to try to normalize the pattern and then to test it against a set of templates to see which it matched. Human recognition, on the other hand, seems to simply disregard changes in size and orientation, as well as breaks in the figure, etc. Although certain perceptual constants do achieve some normalization (apparent size and brightness do not vary as much as corresponding changes in the signal reaching the retina), clearly we do not fully normalize and smooth out the pattern, since we perceive the pattern as skewed, incomplete, large or small, etc., at the same time we recognize it.

More recent programs, rather than normalizing the pattern, seek powerful operators which pick out discriminating traits but are insensitive to distortion and noise. Human pattern recognizers do not employ these artificial expedients either. In those special cases where human pattern recognizers can articulate their cues, these turn out to be not powerful operators which include sloppy patterns and exclude noise, but rather a set of ideal traits which are only approximated in the specific instances of patterns recognized. Distorted patterns are recognized not as falling under some looser and more ingenious set of traits, but as exhibiting the same simple traits as the undistorted figures, along with certain accidental additions or omissions. Similarly, noise is not tested and excluded;

it is ignored as inessential.* Here again we must presuppose the human ability to distinguish the essential from the inessential, which Newell, Shaw, and Simon surreptitiously introduced into their planning program.

Fringe Consciousness. To determine which of a set of already-analyzed patterns a presented pattern most nearly resembles, workers have proposed analyzing the presented pattern for a set of traits by means of a decision tree; or combining the probabilities that each of a set of traits is present, as in Selfridge's Pandaemonium program. Either method uncritically assumes that a human or mechanical pattern recognizer must proceed by a classification based on the analysis of a specific list of traits. It seems self-evident to Selfridge and Neisser that: "A man who abstracts a pattern from a complex of stimuli has essentially classified the possible inputs" [31:238].

Yet, if the pattern is at all complicated and sufficiently similar to many other patterns so that many traits are needed for discrimination, the problem of exponential growth threatens. Supposing that a trait-by-trait analysis is the way any pattern recognizer, human or artificial, must proceed, leads to the assumption that there must be certain crucial traits--if one could only find them, or program the machine to find them for itself--which would make the processing manageable.

---

*Whatever information processing the human brain employs to pick out patterns, this work is no doubt aided by the organization of human receptors. One cannot assume, however, that an organization of the input into perceptual prominences (figure and ground) can be built into the receptors of a digital machine. Such selective receptors would amount to introducing a stage of analogue processing.

Thus one is led to look for a sort of perceptual heuristic, the "powerful operators" which no one as yet has been able to find. And just as the chess masters are not able to provide the programmer with the heuristic short cuts they are supposed to be using, Selfridge and Neisser note in the case of pattern recognition that "very often the basis of classification is unknown, even to [the analyzer]: it is too complex to be specified explicitly" [31:238]. Nevertheless, Selfridge and Neisser assume, like Newell and Simon, that unconsciously a maze is being explored--in this case, that a list of traits is being searched. But the difficulties involved in searching such a list suggest again that not all possibly relevant traits are taken up in series or in parallel and used to make some sort of decision, but that many traits crucial to discrimination are never taken up explicitly at all but remain on the fringe of consciousness.

Moreover, though in chess we are finally reduced to counting out, in perception we need never appeal to <u>any</u> explicit traits. We often recognize an object without recognizing it as one of a type or a member of a class. As Aron Gurwitsch puts it in his analysis of the difference between perceptual and conceptual consciousness:

> Perceived objects appear to us with generic determinations . . . . But--and this is the decisive point--<u>to perceive an object of a certain kind is not at all the same thing as grasping that object as representative or as a particular case of a type</u> [12:203].

Of course, we can sometimes make the cues explicit:

> The first step in the constituting of conceptual consciousness consists in effecting a dissociation within the object perceived in its typicality. The generic traits which until then were immanent and inherent in the perceived thing are detached and disengaged from it. Rendered explicit, these traits can be seized in themselves and crystallize themselves into a new and specific object of consciousness. This object is the concept taken in comprehension. Consequent upon this dissociation, the generic becomes the general. From this aspect it opposes itself to the thing perceived from which it has just been disengaged, and which now is transformed into an example, a particular instance, and, in this sense, into a representative of the concept . . . .
>
> [Thus, cues] can be grasped and become themes [specific traits we are aware of] . . . , whereas previously they only contributed to the constitution of another theme [the pattern] within which they played only a mute role [12:204, 205].

This shift from perceptual to conceptual consciousness (from the perceptive to the mathematical frame of mind, to use Pascal's expression), is not necessarily an improvement. Certain victims of aphasia, studied by Gelb and Goldstein, have lost their capacity for perceptual recognition. All recognition for the patient becomes a question of classification. The patient has to resort to check lists and search procedures, like a digital computer. A typical aphasic can only recognize a figure such as a triangle by listing its traits, i.e., by counting its sides and then thinking: "A triangle has three sides. Therefore, this is a triangle." Such conceptual recognition is time-consuming and unwieldy; the victims of such brain injuries are utterly incapable of getting along in the everyday world.

Evidently, passing from implicit perceptual grouping to explicit conceptual classification--even at some final stage, as in chess--is usually disadvantageous. The fact that we need not conceptualize or thematize the traits common to several instances of the same pattern in order to recognize that pattern, distinguishes human recognition from machine recognition which only occurs on the explicit conceptual level of class membership.

Context-Dependent Ambiguity Reduction. In the cases thus far considered, the traits defining a member of a class, while generally too numerous to be useful in practical recognition, could at least in principle always be made explicit. In some cases, however, such explicitness is not even possible. In recognizing certain complex patterns, as in narrowing down the meaning of words or sentences, the context plays a determining role. The context may simply help us notice those patterns which we can subsequently recognize in isolation. But sometimes an object or person can only be recognized in the context. The unique character of a person's eyes, for example, may depend on the whole face in such a way as to be unrecognizable if viewed through a slit. Moreover, a certain expression of the eyes may bring out a certain curve of the nose which would not be noticed if the nose were in another face; the nose in turn may give a certain twist to the smile which may affect the appearance of the eyes. In such cases, the traits necessary for recognizing these particular eyes cannot be isolated. The context not only brings out the essential features, but is reciprocally determined by them.

In some cases, however, objects recognized as belonging together need not have any traits in common at all. Wittgenstein, in his study of natural language, was led to investigate such cases.

> We see a complicated network of similarities overlapping and criss-crossing: Sometimes overall similarities, sometimes similarities of detail.
>
> I can think of no better expression to characterize these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, color of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way. . . . We extend our concept . . . as in spinning a thread we twist fibre on fibre. And the strength of the thread does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres.
>
> But if someone wishes to say: "There is something common to all these constructions--namely the disjunction of all their common properties"--I should reply: Now you are only playing with words. One might as well say: "Something runs through the whole thread--namely the continuous overlapping of these fibres [42:32].

Those capable of recognizing a member of a "family" need not be able to list <u>any</u> exactly similar traits common to even two members, nor is there any reason to suppose such traits exist. Indeed, formalizing family resemblance in terms of exactly similar traits would eliminate the openness to new cases which is the most striking feature of this form of recognition. No matter what disjunctive list of traits is constructed, one can always invent a new "family" member whose traits are similar to those of the given members

without being exactly similar to any of the traits of any of them.

Here, as in narrowing down the meaning of words or sentences, the context plays a determining role. Recognition of a member of a "family" is made possible not by a list of traits, but by seeing the case in question in terms of its similarity to a paradigm (i.e., typical) case. For example, an unfamiliar painting is recognized as a Cézanne by thinking of a Cézanne we think to be typical. By thinking, if need be, of bridging cases, one can recognize even a deviant case.

Perspicuous Grouping. The above sophisticated but nonetheless very common form of recognition employs a special combination of the three forms of information processing discussed thus far: fringe consciousness, insight, and context dependence. To begin with, the process is implicit. It uses information which remains on the fringes of consciousness.

Seeing the role of insight necessitates distinguishing the generic from the typical, although Gurwitsch uses these two terms interchangeably. Recognition of the generic depends on implicit cues which can always be made explicit. Recognition of the typical, on the other hand, as in the case of family resemblance, depends on cues which cannot be thematized. Recognition of the typical, unlike recognition of the generic, requires insight. A paradigm case serves its function insofar as it is the clearest manifestation of what (essentially) makes all members members of

a given group. Finally, recognition in terms of proximity to the paradigm is a form of context dependence.

Wittgenstein remarks that "a perspicuous representation produces just that understanding which consists in seeing connections" [42:49]. Following Wittgenstein, we will call this combination of fringe consciousness, insight, and context determination "perspicuous grouping." This form of human information processing is an important as the three fundamental forms of information processing from which it is derived.

Conclusion. Human beings are able to recognize patterns under the following increasingly difficult conditions:

1) The pattern may be skewed, incomplete, deformed, and embedded in noise;
2) The traits required for recognition may be "so fine and so numerous" that, even if they could be formalized, a search through a branching list of such traits would soon become unmanageable as new patterns for discrimination were added;
3) The traits may depend upon internal and external context and are thus not isolable into lists;
4) There may be no common traits but a "complicated network of overlapping similarities," capable of assimilating ever new variations.

Any system which can equal human performance, must therefore, be able to:

1) Distinguish the essential from the inessential features of a particular instance of a pattern;
2) Use cues which remain on the fringes of consciousness;

3) Take account of the context;

4) Perceive the individual as typical, i.e., situate the individual with respect to a paradigm case.

Since the recognition of patterns of even moderate complexity may require these four forms of human information processing, work in pattern recognition has not progressed beyond the laborious recognition of a few simple patterns in situations which severely limit variation. It is not surprising, but all the more discouraging, that further progress in game playing, problem solving, and language translation awaits a breakthrough in pattern recognition research.

MISCONCEPTIONS MASKING THE SERIOUSNESS OF CURRENT DIFFICULTIES

The problems facing workers attempting to use computers in the simulation of human intelligent behavior should now be clear. In game playing, the exponential growth of the tree of alternative paths requires a restriction on the paths which can be followed out; in complicated games such as chess, programs cannot select the most promising paths. In problem solving, the issue is not how to direct a selective search, but how to structure the problem so as to begin the search process. In language translation, even the elements to be manipulated are not clear, due to the intrinsic ambiguity of a natural language; in pattern recognition, all three difficulties are inextricably intertwined.

In spite of these grave difficulties, workers in cognitive simulation and artificial intelligence are not discouraged. In fact, they are unqualifiedly optimistic. Underlying their optimism is the conviction that human information processing must proceed by discrete steps like those of a digital computer, and, since nature has produced intelligent behavior with this form of processing, proper programming should be able to elicit such behavior from machines.

The assumption that human and mechanical information processing ultimately involve the same elementary process, is sometimes made naively explicit. Newell, Shaw, and Simon introduce one of their papers with the following remark:

> It can be seen that this approach makes no assumption that the "hardware" of computers and brains are similar, beyond the assumptions that both are general-purpose symbol-manipulating devices, and that the computer can be programmed to execute elementary information processes functionally quite like those executed by the brain [24:9].

They do not even consider the possibility that the brain might process information in an entirely different way than a computer--that information might, for example, be processed globally the way a resistor analogue solves the problem of the minimal path through a network.

In general, workers in cognitive simulation assume that heuristically-guided search techniques reflect the way human beings resolve the difficulties inherent in discrete techniques. Workers in artificial intelligence, although uninterested in copying human information processing techniques, also assume that humans utilize discrete processes--otherwise there would be no reason to expect to find ways to mechanically achieve human results.

Yet judging from their behavior, human beings avoid rather than resolve the difficulties confronting workers in cognitive simulation and artificial intelligence by avoiding the discrete information processing techniques from which these difficulties arise. Why, in the light of this evidence, do those pursuing cognitive simulation assume that the information processes of a computer reveal the hidden information processes of a human being, and why do those working in artificial intelligence assume that there must be a digital way of performing human tasks? Strangely, no one in the field seems to have asked himself these questions.

When intelligent workers are unanimously dogmatic, there must be a reason. Some force in their assumptions must allow them to ignore the need for justification. We must now try to discover why, in the face of increasing difficulties, workers in these fields show such untroubled confidence.

The Associationist Assumption

The development of the high-speed digital computer has strengthened a conviction which was first expressed by Lucretius, later developed in different ways by Descartes and Hume, and finally expressed in nineteenth-century associationist or stimulus-response psychology: thinking must be analyzable into simple determinate operations.*

*The gestaltists claim, in opposition to this school, that thinking involves global processes which cannot be understood in terms of a sequence or even a parallel set of discrete steps. In this context, Newell, Shaw, and Simon's claims to have synthesized the contributions of associationists and gestaltists by, on the one hand, accepting behavioral measures and, on the other, recognizing that "a human being

The suitably programmed computer can be viewed as a working model of the mechanism presupposed by this theory. Artificial intelligence has in this way made associationism operational and given it a second wind.

The affinity between this venerable but somewhat outdated conception of mental processes and the presuppositions of workers in artificial intelligence is often quite explicit. As Lindsay says in his article on "Machines which Understand Natural Language,"

> A list structure is a form of associative memory, wherein each symbol is tagged by an indicator which tells the machine the location of a related symbol. So far this corresponds to the associative bonds which are the basic concept of stimulus-response psychology [14:221].

Early success in artificial intelligence has so strengthened this associationist assumption that no one feels called upon to defend associationism in the face of mounting evidence in both experimental psychology and in the artificial intelligence field itself that, although machines do, people do not perform intelligent tasks by simple determinate steps. To determine whether the confidence exhibited by workers in cognitive simulation and artificial intelligence is justified, we must evaluate the empirical and philosophical arguments offered for associationism.

---

is a tremendously complex, organized system" [26:280,293] shows either a will to obscure the issues or a total misunderstanding of the contribution of each of these schools.

Empirical Evidence for the Associationist Assumption: Critique of the Scientific Methodology of Cognitive Simulation. The empirical justification of the associationist assumption poses a question of scientific methodology--the problem of the evaluation of evidence. Gross similarities of behavior between computers and people do not justify the associationist assumption, nor does the present inability to demonstrate these similarities alone justify its rejection. A test of the associationist assumption requires a detailed comparison of the steps involved in human and machine information processing. Newell, Shaw, and Simon conscientiously note the similarities and differences between human protocols and machine traces recorded during the solution of the same problem. We must now turn to their evaluation of the evidence thus obtained.

After carefully noting the exceptions to their program, Newell and Simon conclude that their work

> provide[s] a general framework for understanding problem-solving behavior . . . and finally reveals with great clarity that free behavior of a reasonably intelligent human can be understood as the product of a complex but finite and determinate set of [presumably associationist] laws [26:293].

This is a strangely unscientific conclusion to draw from a program which "provides a complete explanation of the subject's task behavior with five exceptions of varying degrees of seriousness" [26:292]. For Newell and Simon acknowledge that their specific theories--like any scientific theories--must stand or fall on the basis of their

generality, that is, the range of phenomena which can be explained by the programs [24:9].

There seems to be some confusion concerning the universality of scientific laws. Scientific laws do not admit of exceptions, yet here the exceptions are honestly noted--as if the frank recognition of these exceptions mitigates their importance, as if Galileo might, for example, have presented the law of falling bodies as holding for all but five objects which were found to fall at a different rate. Not that a scientific theory must necessarily be discarded in the face of a few exceptions; there are scientifically sanctioned ways of dealing with such difficulties. One can, to begin with, hold on to the generalization as a working hypothesis and wait to announce a scientific law until the exceptions are incorporated. A working hypothesis need not explain <u>all</u> the data. When, however, one claims to present a theory, let alone a "general framework for understanding," then this theory must account for <u>all</u> the phenomena it claims to cover--either by subsuming them under the theory or by showing how, according to the theory, one would expect such exceptions.

Even without exceptions, the theory would not be general, since the available evidence has necessarily been restricted to those most favorable cases where the subject can to some extent articulate his information processing protocols (game playing and the solution of simple problems as opposed to pattern recognition and the acquisition and use of natural language). But even if we were to ignore this difficulty and require only a special theory of problem solving, ordinary scientific standards of accounting for exceptions

would invalidate all cognitive simulation theories so far presented. As things stand, even after ad hoc adjusting of the program to bring it into line with the protocol--itself a dubious procedure--a machine trace never completely matches the protocol and the exceptions, while carefully noted, are never explained.

There is one other acceptable way of dealing with exceptions. If one knew, on independent grounds, that mental processes must be the product of discrete operations, then exceptions could be dealt with as accidental difficulties in the experimental technique, or challenging cases still to be subsummed under the law. Only then would those involved in the field have a right to call each program which simulated intelligent behavior--no matter how approximately--an achievement and to consider all set-backs nothing but challenges for sharper heuristic hunting and further programming ingenuity. The problem, then, is how to justify independently the associationist assumption that all human information processing proceeds by discrete steps. Otherwise the exceptions along with the narrow range of application of the programs and the lack of progress during the last few years, tend to deconfirm, rather than confirm, the hypothesis. The "justification" seems to have two stages.

In the early literature, instead of attempting to justify this important and questionable assumption, Newell, Shaw, and Simon present it as a postulate, a working hypothesis which directs their investigation. "We postulate that the subject's behavior is governed by a program organized from a set of elementary information processes" [24:9]. This postulate, which alone might seem rather arbitrary, is in

turn sanctioned by the basic methodological principle of parsimony. This principle enjoins us to assume tentatively the most simple hypothesis, in this case that all information processing resembles that sort of processing which can be programmed on a digital computer. We can suppose, for example, that in chess, when our subject is zeroing in, he is unconsciously counting out. In general, whenever the machine trace shows steps which the subject did not report, the principle of parsimony allows us to suppose that the subject unconsciously performed these steps. So far this is perfectly normal. The principle of parsimony justifies picking a simple working hypothesis as a guide to experimentation. But of course the investigations must support the working hypothesis; otherwise it must eventually be discarded.

The divergence of the protocols from the machine trace, as well as the difficulties raised by planning, indicate that things are not so simple as our craving for parsimony leads us to hope. In the light of these difficulties, it would be natural to revise the working hypothesis, just as scientists had to give up the Bohr conception of the atom; but at this point, research in cognitive simulation deviates from acceptable scientific procedures. In a recent publication, Newell and Simon announce:

> There is a growing body of evidence that the elementary information processes used by the human brain in thinking are highly similar to a subset of the elementary information processes that are incorporated in the instruction codes of the present-day computers [35:282].

What is this growing body of evidence? Have the gaps in the protocols been filled and the exceptions explained? Not at all. The growing body of evidence seems to be the very programs whose lack of universality would cast doubt on the whole project but for the independent assumption of the associationist hypothesis. The associationist assumption must have at first been taken as independently justified, since the specific programs are presented as established theories, and yet now the assumption is recognized as an hypothesis whose sole confirmation rests on the success of the specific programs.

An hypothesis based on a methodological principle is often confirmed later by the facts. What is unusual and inadmissible is that, in this case, the hypothesis produces the evidence by which it is later confirmed. Thus, no empirical evidence exists for the associationist assumption. In fact, the supposed empirical evidence presented for the assumption tends, when considered in itself, to show that the assumption is empirically untenable.

This particular form of methodological confusion is restricted to those working in cognitive simulation, but even workers in artificial intelligence share their belief in the soundness of heuristic programs, their tendency to think of all difficulties as accidental, and their refusal to consider any set-backs as disconfirming evidence. Concluding from the small area in which search procedures are partially successful, workers in both fields find it perfectly clear that the unknown and troublesome areas are of exactly the same sort. Thus, all workers proceed as if the credit of the associationist assumption were assured,

even if all do not--like those in cognitive simulation--attempt to underwrite the credit with a loan for which it served as collateral. For workers in the field, the associationist assumption is not an empirical hypothesis which can be supported or disconfirmed, but some sort of philosophical axiom whose truth is assured a priori.

A Priori Arguments for the Associationist Assumption: Conceptual Confusions Underlying Confidence in Artificial Intelligence. As stated in artificial intelligence literature, the claim that all human information processing can in principle be simulated or at least approximated on a digital computer, presupposes the validity of the associationist assumption. Feigenbaum, for example, asserts:

> . . . Human thinking is wholly information-processing activity within the human nervous system; these information processes are perfectly explicable; . . . digital computers, being general information-processing devices, can be programmed to carry out any and all of the information processes thus explicated [6:248,249].

The statement that a computer is a general information-processing device does indeed imply that a digital computer can process any information which is completely formalized, i.e., expressed in exhaustive and unambiguous form. But this is significant for work in artificial intelligence only if information processes in humans are also "perfectly explicable," i.e., reducible to discrete operations. Feigenbaum gives no argument to back up his claim.

Such an assertion, however, is by no means obvious. If it is supposed to gain plausibility from the physiological fact that the human nervous system operates with all-or-none switches like a digital computer, it is antiquated by the recent discoveries in brain physiology.

> . . . In the higher invertebrates we encounter for the first time phenomena such as the graded synaptic potential, which before any post synaptic impulse has arisen can algebraically add the several incoming presynaptic barrages in a complex way. These incoming barrages are of different value depending upon the pathway and a standing bias. Indeed, so much can be done by means of this graded and nonlinear local phenomenon prior to the initiation of any post-synaptic impulse that we can no more think of the typical synapse in integrative systems as being a digital device exclusively as was commonly assumed a few years ago, but rather as being a complex analog device . . . [4:172].

If this assertion (that human information processing is explicable in discrete terms) claims to be based on a description of human experience and behavior, it is even more untenable. Certain forms of human experience and behavior clearly require that some of the information being processed not be made perfectly explicit. Consider a specific example from gestalt psychology: When presented with the equal line segments in the Muller-Lyer illusion (Fig. 1), the subject cannot help but see the upper line as shorter than the lower. The lines at the end of each segment (which are not considered explicitly,

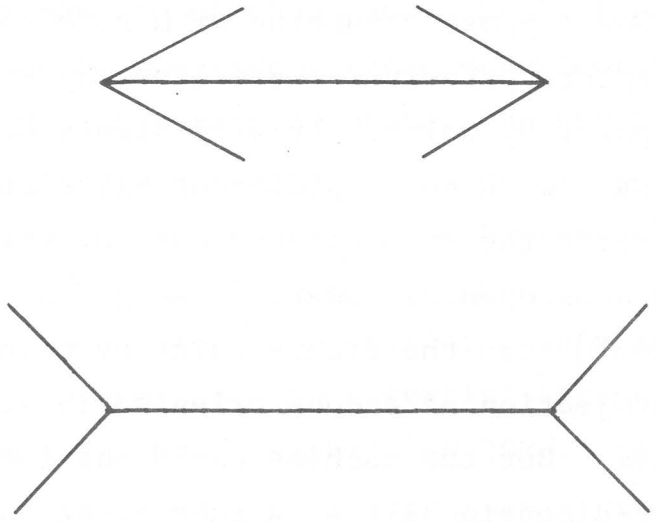


Fig. 1--Muller-Lyer Illusion

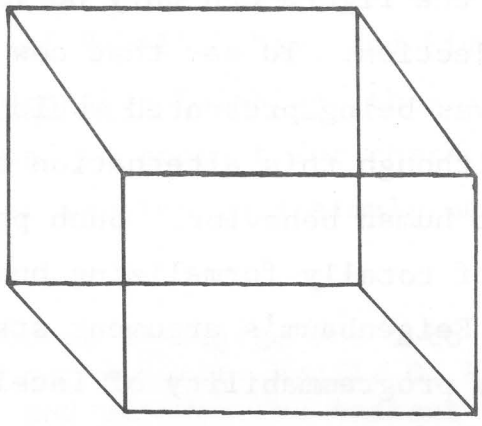


Fig. 2--Necker Cube

but which rest on the fringes of the perceptual field) affect the appearance of the lines on which attention is centered. Now suppose a machine with some sort of electronic perceptors perceives these lines by scanning them explicitly point by point. It will simply perceive the lines as equal, with no suspicion of illusion.

Or consider the even clearer case of the Necker Cube (Fig. 2) seen as opening toward or away from the viewer. A machine could scan the figure point by point and analyze it as the projection of a cube oriented in either of two possible ways. But the machine could not interpret the figure three-dimensionally as a cube first in one, then in the other of these orientations. Such an interpretation would require the machine to focus on certain aspects of the figure while leaving others in the background, and the machine lacks precisely this figure-ground form of representation. For it, every point of the figure is equally explicit; thus the figure can only be interpreted as an ambiguous flat projection. To say that now one, now the other orientation was being presented would make no sense in such a program, although this alternation of perspectives could easily affect human behavior. Such phenomena challenge the possibility of totally formalizing human information processing. As Feigenbaum's argument stands, the case for the necessary programmability of intelligent human behavior has not been made.

Still, the associationist assumption is not so easily dismissed. After all, a device does exist which can detect the Muller-Lyer illusion and respond to the difference

between the two aspects of the cube: the human brain.* And if this device obeys the laws of physics and chemistry, which we have every reason to suppose it does, then we ought to be able to build an analogous device which might, for example, take the form of an analogue computer using ion solutions whose electrical properties change with various local saturations.†

Further, knowing the solutions and how they work enables us at least in principle to write the physico-chemical equations describing such wet components and to solve these equations on a dry digital computer. Thus, given enough memory and time, any computer--even such an analogue computer--could be simulated on a digital computer. In general, by accepting the fundamental assumptions that the brain is part of the physical world and that all physical processes can be described in a mathematical formalism which can in turn be manipulated by a digital computer, one can arrive at the strong claim that all human information processing, whether formalizable or not, can be carried out on a digital machine.

---

*It seems self-evident that we could simulate intelligent behavior if we could build or simulate a device which functioned exactly like the human brain. But even this could be challenged if it could be shown that the body plays a crucial role in making possible intelligent behavior. This view has been developed by Maurice Merleau-Ponty in his book Phenomenology of Perception, and may be implicit in the work of MacKay, but will not be defended in this Paper.

†MacKay seriously considers such a possibility: "It may well be that only a special-purpose 'analogue' mechanism could meet all detailed needs . . . . We on the circuit side had better be very cautious before we insist that the kind of information processing that a brain does can be replicated in a realizable circuit. Some kind of 'wet' engineering may turn out to be inevitable" [16:16].

This claim may well account for the simulators' smugness, but what in fact is justified by the fundamental truth that every form of information processing (even those in which in practice can only be carried out on an analogue computer) must in principle be simulable on a digital computer? Does it really prove the associationist claim that, even when a human being is unaware of using discrete operations in processing information, he must nonetheless be carrying on unconscious searching, sorting, and storing?

Consider again the ion solution which might duplicate the information processing in the Muller-Lyer illusion. Does the solution, in reaching equilibrium, go through the series of discrete steps a digital computer would follow in solving the equations which describe this process? In that case, the solution is solving in moments a problem which it would take a machine centuries to solve--if the machine could solve it at all. Is the solution an ultra-rapid computer, or has it got some secret, clever heuristic like the chess master, which simplifies the problem? Obviously, neither. The fact that we can describe the process of reaching equilibrium in terms of equations and then break up these equations into discrete elements in order to solve them on a computer does not show that equilibrium is actually reached in discrete steps. Likewise, we need not conclude from the claim that all continuous processes involved in human information processing can be formalized and calculated out discretely, that any discrete processes are actually taking place. Once the a priori argument for associationism, based on the all-purpose character of the digital computer, is

restated so as to be defensible, it turns out not to be an argument for associationism at all.

CONCLUSION

Without the associationist assumption to fall back on, what encouragement can workers in cognitive simulation and artificial intelligence draw from the argument that, even though the brain does not process information in discrete operations, we can simulate the brain on a digital computer and thus by discrete operations produce the same results the brain produces?

To begin with, what would such a computer program tell us about operations on the information-processing level? According to Newell, Shaw, and Simon, a description of operations on the information-processing level is a theory in psychology and not physiology. Psychological operations must be the sort which human beings at least sometimes consciously perform in processing information--e.g., searching, sorting, and storing--and not physico-chemical processes in the organism. Thus a chess player's report as he zeroed in on his Rook, "And now my brain reaches the following chemical equilibrium, described by the following array of differential equations," would describe physiological processes no doubt correlated with information processing, but not that information processing itself.

Similarly, one must delimit what can count as information processing in a computer. A digital computer solving the equations describing an analogue information-processing device and this simulating its <u>function</u> is not thereby simulating its <u>information processing</u>. It is not processing

the information which is processed by the simulated analogue, but entirely different information concerning the physical or chemical properties of the analogue. Thus the strong claim that every processable form of information can be processed by a digital computer is misleading. One can only show that, for any given type of information, a digital computer can in principle be programmed to simulate a device which can process that information. This does not support Feigenbaum's assertion that human information processes are perfectly explicable, and therefore fails to show that "digital computers being general information-processing devices, they can be programmed to carry out any and all information processes."

Confidence of progress in cognitive simulation is thus as unfounded as the associationist assumption, but this realization leaves untouched the weaker claim of workers in artificial intelligence that human intelligent behavior--not human information processing--can be simulated by using digital computers. Nothing that has been said thus far suggests that digital computers could not process, in their own way, the information which human beings process. Indeed, at first sight, our results might seem encouraging for work in artificial intelligence. We have seen that, in principle, a digital computer can simulate any physical information processing system. In fact, however, no comfort can be gained from this substitute for the associationist assumption, since this "principle" cannot be realized in practice. We do not know the equations describing the physical processes in the brain, and even if we did, the solution of the equations describing the simplest reaction would take a prohibitive amount of time.

The facts that the associationist assumption cannot be defended on empirical or on a priori grounds, and that the simulation of the brain is in practice impossible, do not show that the task set for artificial intelligence is hopeless. However, they eliminate the only argument which suggests any particular reason for hope. The associationist assumption asserted that human and mechanical information processing proceed by discrete operations, leaving artificial intelligence the promising task of finding them. Without the defense provided by the associationist assumption, all the difficulties of artificial intelligence during the past few years take on new significance: there is no reason to deny the evidence that human and mechanical information processing proceed in entirely different ways. At best, research in artificial intelligence can write programs which allow the digital machine to approximate, by means of discrete operations, the results which human beings achieve by avoiding rather than resolving the difficulties inherent in discrete techniques.

Is such research realistic? Can one introduce search heuristics which enable the speed and accuracy of computers to bludgeon through in those areas where human beings use more elegant techniques? Lacking any a priori basis for confidence, we can only turn to the empirical results obtained thus far. That brute force can succeed to some extent is demonstrated by the early work in the field. The present difficulties in game playing, problem solving, language translation, and pattern recognition, however, indicate a limit to our ability to substitute one kind of information processing for another. Only experimentation

can determine the extent to which newer and faster machines, better programming languages, and clever heuristics can continue to push back the frontier. Nonetheless, the dramatic slowdown in the fields we have considered and the general failure to fulfill earlier predictions suggest the boundary may be near.

Part III

THE FUTURE OF ARTIFICIAL INTELLIGENCE

No valid empirical or *a priori* arguments have been put forward to support the associationist assumption, and therefore there is no reason to expect continuing progress in artificial intelligence. On the other hand, no arguments have been put forward to deny the possibility of such progress. Are there any reasons for denying that such continuing progress is possible? That is, are there any reasons to suppose that the unexpected difficulties which have appeared in all areas of artificial intelligence research indicate a necessary limit to what can be accomplished with digital computers in this field?

To understand these difficulties and show that they are more than temporary, we would have to show that mechanical information processing has inherent limitations from which human information processing is free. We have already considered the processing itself; here there is no way to fix a limit to the degree of approximation clever heuristics might achieve. We have not yet considered the information to be processed. I propose to show now that, near or far off, there does exist a boundary to possible progress in the field of artificial intelligence: given the nature of the information to be processed, the contribution of the uniquely human forms of information processing which we have considered are indispensable, since they alone provide access to information inaccessible to a mechanical system.

THREE NON-PROGRAMMABLE FORMS OF INFORMATION

Machines are perfect Cartesians. They are able to deal only with the determinate and discrete bits of information which Descartes called "clear and distinct ideas." Newell describes GPS as "a program for accepting a task environment defined in terms of discrete objects" [20:17]; Feigenbaum and Feldman extend this basic requirement when they assert that the only constraint on the computer user "is that his statements be unambiguous and complete" [8:271]. They, like Descartes, consider this "a blessing rather than a limitation, for it forces a refreshing rigor on builders of models of human thought processes." This may well be true for cognitive simulation considered as a branch of psychology, but it ignores the more general attempt to introduce mechanical information processing into all areas of intelligent activity which are now the exclusive province of human beings. Simon predicts that:

> There will be more and more applications of machines to take the place of humans in solving ill-structured problems; just as machines are now being more and more used to solve well-structured problems [32:8].

If the machine is only able to handle unambiguous, completely structured information, however, how can it deal with the ill-structured data of daily life? Indeed, here the project of using digital computers to simulate or even approximate human information processing seems to reach its absolute

limit; the computer cannot be given the information it is to process.*

This limit is manifest in each of the areas in which a uniquely human form of information processing is necessary to avoid the difficulties faced by digital computers. In these areas, if we restrict ourselves to information which can be fed to digital computers and yet try to write a program which rivals everyday human information processing, a contradiction develops within the program itself.

The Infinity of Facts and the Threat of Infinite Progression

In the area of game playing, as we have seen, the array of branching possibilities to be searched may become so large that heuristics are necessary to eliminate a certain number of possible alternatives. These heuristics save the day by pruning the search tree, but they also discard some combinations a human player could consider, so situations will always occur in which the machine cannot pursue the chain of moves which contains the winning combination; thus, there will always be games that people can win and machines cannot.

---

*In The Process of Creative Thinking, Newell, Shaw, and Simon list four characteristics of creative thought, the fourth of which is: "The problem as initially posed was vague and ill defined, so that part of the task was to formulate the problem itself" [21:4]. They claim that, "a problem-solving process [presumably their own] can exhibit all of these characteristics to a greater or lesser degree . . ." [24:4]. In the light of Newell's statement that "GPS is a program for accepting a task environment defined in terms of discrete objects . . ." [20:17], one can only wonder whether, in the literature of artificial intelligence, zero counts as a lesser degree.

Concerning formal finite games like chess, this is only a practical objection. In principle, at least, the whole maze could be calculated out; or one could introduce a random element, as Newell once suggested which, while complicating the program without improving the play, would answer the objection that there were specific moves whose consideration was forbidden to the machine.

However, in a non-formal game like playing the horses--which is still much more systematic than the everyday ill-structured problems that Simon predicted machines would be able to handle--an unlimited set of conditions become relevant. In placing a bet, we can usually restrict ourselves to facts about the horse's age, jockey, and past performance--and perhaps, restricted to these, the machine could do fairly well, perhaps better than an average handicapper--but there are always other factors, such as whether the horse is allergic to goldenrod or whether the jockey has just had a fight with the owner--which <u>may</u> in some cases be decisive. These possibilities remain on the fringes of consciousness. If the machine were to examine explicitly each of these possibly relevant factors as determinate bits of information, in order to determine whether to take it into consideration or ignore it, it could never complete the calculations necessary to predict the outcome of a single race. If, on the other hand, the machine systematically excluded possibly relevant factors in order to complete its calculations, then the machine would sometimes be incapable of performing as well as an intelligent human.

Descartes, who was the first to ask whether a machine could imitate all the actions of men, comes to a similar conclusion.

> . . . Although such machines could do many things as well as, or perhaps even better than, men, they would infallibly fail in certain others . . . . For while reason is a universal instrument which can be used in all sorts of situations, the organs [of a machine] have to be arranged in a particular way for each particular action. From this it follows that it is morally impossible that there should be enough different devices [i.e., states] in a machine to make it behave in all the occurrences of life as our reason makes us behave [5:36].

Even the appeal to a random element will not help here, since to be able to take up a sample of excluded possibilities at random so that no possibility is in principle excluded, the machine would have to be explicitly provided with a list of all such other possibly relevant facts or a specific set of routines for exploring all classes of possibly relevant facts, so that no facts were in principle inaccessible. This is just what could be done in a completely defined system such as chess, where a finite number of concepts determines totally and unequivocally the set of all possible combinations in the domain, but in the real world the list of such possibly relevant facts, or even possibly relevant classes of facts, is indefinitely large ("infinite in a pregnant sense," to use Bar-Hillel's phrase), and cannot be exhaustively listed. The ability to retain this infinity of facts on the fringes of consciousness allows human beings access to the open-ended information characteristic of everyday experience, without leading to the inconsistency of requiring an incompletable series of data-gathering operations before the data processing can begin.

The Indeterminacy of Needs and the Threat of Infinite Regress

In problem solving, the contradiction takes a different form. If, using only digital programs, we try to process the ill-structured data in which real-life problems are posed, we face an infinite regress.

A problem can in principle always be solved on a digital computer, provided the data and the rules of transformation are explicit. However, Newell, Shaw, and Simon have pointed out that--even in the case of simple logic problems--finding a path through the maze of possible combinations requires a planning program. In the case of formal problems, planning is a matter of practical necessity; in the case of ill-defined problems, it is necessary in principle. Since an indefinite amount of data may be relevant for the solution of an ill-defined problem, one cannot even in principle try all the permutations of the possibly relevant data in seeking a solution. Thus, one needs to structure the problem, to determine both which facts from the environment are relevant, and which operations bring about essential transformations.

According to Minsky, Simon's group working on GPS has set itself the goal of giving the problem-solving program the problem of improving its own operation [10:117]. This, one might hope, would enable a computer to discover the data and operations essential to the solution of a certain type of problem and write a plan for solving the problem. But a difficulty immediately arises: such a planning program itself would require a distinction between essential

and inessential operators. Unless, at some stage, the programmer himself introduces this distinction, he will be forced into an infinite regress of planning programs, each one of which will require a higher-order program to structure its ill-structured data.

The nature of the essential/inessential distinction itself explains this regress. Newell, Shaw, and Simon remark that the distinction they introduce in setting up their planning program is pragmatic. Such a pragmatic distinction is made in terms of goals. These goals in turn are determined by needs, and these needs are not themselves always precise. Some needs are first experienced as an indeterminate sense that the present situation is unsatisfactory; we can determine these needs only when we discover what action reduces this uneasiness. Thus, needs and goals cannot be introduced as determinate data which can then be used in structuring the problem. Often only in structuring or solving the problem do they become precise. Only the uniquely human form of information processing which uses the indeterminate sense of dissatisfaction to pragmatically structure ill-structured problems enables us to avoid the problem-solving regress.

The Reciprocity of Context and the Threat of Circularity

The meaning of a word is determined by its context, but also contributes to the meaning of that context. As long as all the meanings in question are left somewhat ambiguous (i.e., as long as possible ambiguities are not resolved in advance, and the meanings are made only as determinate as necessary for the particular activity in

question), there is no contradiction in this notion of the totality of elements determining the significance of each one. If, however (in order to describe the situation in language suited to a computer), we try to explicate the meaning of a word used in a context, then we find ourselves obliged to resolve all the ambiguities in the context. Since the meaning of each term contributes to the meaning of the context, every word must be made determinate before any word can be made determinate, and we find ourselves involved in a circle.

This situation may even arise in a completely formal system if we try to use heuristics to avoid exponential growth. In developing a heuristic program for playing chess, one must evaluate the positions arrived at. This evaluation must depend on the evaluation of parameters, which measure success in achieving certain goals. To evaluate these parameters, one must assume that any parameter can be considered independently of the others. For example, in explaining the evaluation of "Material Balance," Newell, Shaw, and Simon note that: "For each exchange square a static exchange value is computed by playing out the exchange with all the attackers and defenders assuming no indirect consequences like pins, discovered attacks, etc." [22:59] (italics added).

Newell, Shaw, and Simon seem to assume that such specification, independent of the other parameters, is simply a matter of caution and ingenuity. Feigenbaum and Feldman make the same assumption when they casually remark that, "before the . . . chess model . . . could be programmed, the meaning of the words 'check' and

'danger' would have to be specified" [8:271]. What counts as "danger," however, depends not simply on whether a piece is attacked, but whether it is defended by a less valuable piece; whether, in the capturing move, a check is revealed or a forced mate is allowed. In the case of a trade, it further depends on who is ahead, the stage of the game, who is on the offensive, who has the tempo, etc. Clearly, for a more and more refined definition of danger, a larger segment of the total situation will have to be considered. Moreover, at some point the factors to be taken into account, such as tempo or possibility of a forced mate, will themselves have to be defined in terms which involve the determination of whether pieces are in danger.

It is not clear how complete such a definition would have to be for a heuristic program to be able to play good or even mediocre chess. The poor performance of chess programs may indicate that thus far evaluations have been too static and crude. Perhaps, as Newell remarks in his discussion of the difficulties in problem-solving programs, "something has been assumed fixed in order to get on with the program, and the concealed limitation finally shows itself" [19:56]. If, however, one attempts to refine the evaluation of parameters, the interdependence of such definitions will eventually be revealed by a loop, which could be eliminated only by sacrificing the flexibility of the definitions of the parameters involved. At this point, the limits of a heuristic chess program will have become a matter of principle rather than simply of practice.

The reason a human player does not go into a corresponding loop is that his definitions are neither completely flexible and sophisticated--so as to take into account all possible situations--nor are they static and crude. His definitions are adjustable. Thus he is able, for example, to define "danger" as precisely as necessary to make whatever decision the situation requires, without at the same time being obliged to try to eliminate all possible ambiguity. The human player never need make the whole context explicit in working out any particular move.

The digital computer by definition lacks this ambiguity tolerance. A program for collecting information concerning parameters must either arbitrarily isolate the area in question and restrict the definition of the parameters, or take into account all consequences, no matter how indirect. In the first case, the machine's play will be crude; in the second, the program will contain a loop and the machine will not be able to play at all.

Newell, in his thoughtful paper on the problems involved in program organization, seems on the verge of recognizing the importance of the flexibility inherent in human information processing. He remarks that "sequential processing . . . built into the basic structure of our machines . . . encourages us to envision isolated processes devoted to specific functions, each passively waiting in line to operate when its turn comes" [19:10], and he notes that "it seems a peculiar intelligence which can only reveal its intellectual powers in a fixed pattern" [19:18]. Yet Newell is still convinced that more ingenious programs or the substitution of parallel for sequential processing can

remove these difficulties. He does not seem to realize that, if one attempts to use a computer which can only deal with discrete unambiguous information to process context-dependent information, the isolation of processes is necessary if one is to avoid circularity.

Only Shannon seems to be aware of the true dimensions of the problem: that by its very nature as a discrete machine, a digital computer cannot cope with intrinsic ambiguity. In a discussion of "What Computers Should Be Doing," he observes that:

> . . . Efficient machines for such problems as pattern recognition, language translation, and so on, may require a different type of computer than any we have today. It is my feeling that this computer will be so organized that single components do not carry out simple, easily described functions. . . . Can we design . . . a computer whose natural operation is in terms of patterns, concepts, and vague similarities, rather than sequential operations on ten-digit numbers? [10:309-310]

AREAS OF INTELLIGENT ACTIVITY CLASSIFIED WITH RESPECT TO THE POSSIBILITY OF ARTIFICIAL INTELLIGENCE IN EACH

This section discusses the various areas of intelligent activity which have been or might be attacked by workers in artificial intelligence, in order to determine to what extent intelligent activity in each area presupposes the three uniquely human forms of information processing. We can thus account for what success has been attained and predict what further progress can be expected. There are four distinct areas of intelligent behavior (cf. Table 1). The first and third are adaptable

Table 1

CLASSIFICATION OF INTELLIGENT ACTIVITIES

I. Associationistic	II. Non-formal	III. Simple Formal	IV. Complex Formal
Characteristics of Activity			
Irrelevance of meaning and context.	Dependent on meaning and context, which are not explicit.	Meanings completely explicit and context-independent.	In principle, same as III; in practice, internally context-dependent, independent of external context.
Learned by repetition.	Learned by perspicuous examples.	Learned by rule (exception: checkers).	Learned by rule and practice.
Field of Activity (and Appropriate Procedure)			
Memory games, e.g., "Geography" (association).	Ill-defined games, e.g., riddles (perceptive guess).	Computable or quasi-computable games, e.g., nim or checkers (seek algorithm or count out).	Uncomputable games, e.g., chess or go (global intuition and detailed counting out).
Maze problems (trial and error).	Structurable problems (insight).	Combinatory problems (non-heuristic means/ends analysis).	Complex combinatory problems (planning and maze calculation).
Word-by-word translation (mechanical dictionary).	Translating a natural language (understanding in context of use).	Proof of theorems in decidable math (seek algorithm).	Proof of theorems in undecidable math (intuition and calculation).
Instinctive recognition of rigid patterns (conditioned response).	Recognition of varied and distorted patterns (recognition of generic or use of paradigm case).	Recognition of simple rigid patterns, e.g., reading typed page (search for traits whose conjunction defines class membership).	Recognition of complex patterns in noise (search for regularities).
Kinds of Program			
Decision tree, list search.	None.	Algorithm or limit on growth of search tree.	Search-pruning heuristics.

to digital computer simulation, while the second is totally intractable, and the fourth is amenable to only a small extent. The assumption that all intelligent behavior can be mapped on a multi-dimensional continuum has encouraged workers to generalize from success in the two promising areas to unfounded expectations of success in the other two.

Area I includes all forms of elementary associationistic behavior where meaning and context are irrelevant to the activity concerned. Learning nonsense syllables is the most perfect example of such behavior so far programmed, although any form of conditioned reflex would serve as well. Also some games, such as the game sometimes called "geography" (which simply consists of finding a country whose name begins with the last letter of the previously named country), belong in this area. In language translating, this is the level of the mechanical dictionary; in problem solving, that of pure trial-and-error routines.

Area II might be called the area of non-formal behavior. This includes all our everyday activities in indeterminate situations. The most striking example of this controlled imprecision is our use of natural languages. This area also includes games in which the rules are not definite, such as guessing riddles. Pattern recognition in this domain is based on recognition of the generic or typical, and the use of the paradigm case. Problems on this level are unstructured, requiring a determination of what is relevant and insight into which operations are essential, before the problem can be attacked.* Techniques on this

*The activities found in Area II can be thought of as the sort of "milestones" asked for by Paul Armer in his article, "Attitudes toward Intelligent Machines": "A clearly

level are usually taught by example and followed intuitively without appeal to rules. We might adopt Pascal's terminology and call Area II the home of the esprit de finesse.

Area III on the other hand, is the domain of the esprit de géométrie. It encompasses the conceptual rather than the perceptual world. Problems are completely formalized and completely calculable. For this reason, it might best be called the area of the simple-formal.

In Area III, natural language is converted into formal language, of which the best example is logic. Games have precise rules and can be calculated out completely, as in the case of nim or tic-tac-toe, or at least sufficiently to dispense with search-pruning heuristics (checkers). Pattern recognition on this level takes place according to determinate types, which are defined by a list of traits characterizing the individuals which belong to the class in question. Problem solving takes the form of reducing the distance between means and ends by recursive application of formal rules. The formal systems in this area, as we have defined it, are characteristically simple enough to be manipulated by algorithms which require no search procedure at all (for example, Wang's logic program), or require search-limiting but not search-pruning procedures (Samuel's checker program). Heuristics are not only unnecessary here, they are a positive handicap, as the

---

defined task is required which is, at present, in the exclusive domain of humans (and therefore incontestably 'thinking') but which may eventually yield to accomplishment by machines" [1:397]. We contend that such machines could not be digital computers; they would have to exhibit the sort of flexibility suggested by Shannon.

relative success of the NSS and the Wang logic programs have strikingly demonstrated. In this area, artificial intelligence has had its only unqualified successes.

Area IV, complex-formal systems, is the most difficult to define and has generated most of the misunderstandings and difficulties in the field. The difference between the simple-formal and the complex-formal systems need not be absolute. As used here, "complex-formal" includes systems in which exhaustive computation is impossible (undecidable domains of mathematics) as well as systems which, in practice, cannot be dealt with by exhaustive enumeration (chess, go, etc.).*

---

*It is difficult to classify and evaluate the various one-purpose programs that have been developed for motor design, line balancing, integrating, etc. They are not clearly successful programs, until a) like the chess and checker programs they are tested against human professionals; and b) the problems attacked by these programs have, if possible, been formalized so that these heuristic programs can be compared with non-heuristic programs designed for the same purpose. (Wherever such a comparison has been made--in checkers, logic, pattern recognition, chess--the non-heuristic programs have proved either equal or superior to their heuristic counterparts.)

Programs which simulate investment banking procedures and the like have no bearing on cognitive simulation or artificial intelligence. They merely show that certain forms of human activity are sufficiently simple and stereotyped to be formalized. Intelligence was surely involved in <u>formulating</u> the rules which investors now follow in making up a portfolio of stocks, but the <u>formalization</u> of these rules only reveals them to be explicable and unambiguous, and casts no light on the intelligence involved in discovering them or in their judicious application. The challenge for artificial intelligence does not lie in such <u>ex post facto</u> formalization of a specific task, but

.The literature of artificial intelligence generally fails to distinguish these four areas. For example, Newell, Shaw, and Simon announce that their logic theorist "was devised to lean how it is possible to solve difficult problems such as proving mathematical theorems [III or IV], discovering scientific laws from data [II and IV], playing chess [IV], or understanding the meaning of English prose [II]" [24:109]. This confusion has two dangerous consequences. First there is the tendency to think that heuristics discovered in one field of intelligent activity, such as theorem proving, must tell us something about the information processing in another area, such as the understanding of a natural language. Thus, certain simple forms of information processing applicable to Areas I and III are imposed on Area II, while the unique forms of information processing in this area are overlooked.

Second there is the converse danger that the informal processes used in Area II may be covertly introduced in the programs for dealing with other areas, particularly Area IV, with even more disastrous consequences. The success of artificial intelligence in Area III depends upon avoiding anything but discrete and determinate operations. The fact that, like the simple systems in Area III, the complex systems in Area IV are formalizable, leads the

---

rather in Area II in which behavior is flexible and not strictly formalizable, in Area III where the formalization is sufficiently complex to require elegant techniques in order to reach a solution, and in Area IV where the formal system is so complex that no decision procedure exists and one has to resort to heuristics.

simulator to suppose the intelligent activities in Area IV are likewise amenable to programming on a digital computer. The difference in degree between simple and complex systems, however, turns out in practice to be a difference in kind; exponential growth becomes a serious problem. When he discovers his inability to cope with the problems of complex-formal systems, using the techniques which worked with simple-formal systems, the programmer (unaware of the differences between the four areas) may inconsistently introduce procedures borrowed from the observation of behavior in Area II--e.g., evaluation of position in chess, planning in problem solving. These procedures are useful only in conjunction with one or more of the specifically human forms of information processing--a heuristic chess program, using context-dependent evaluations, presupposes ambiguity tolerance; the introduction of planning into simple means-end analysis presupposes a distinction between essential and inessential operations, etc. The programmer, of course, does not suspect that he is treating the formal system in Area IV as if it were a non-formal system, but in fact he is introducing into the continuity between Areas III and IV a discontinuity similar to the discontinuity between Areas I and II. Thus, problems which in principle should only arise in trying to program the ill-structured and open-ended activities of daily life, arise in practice for complex-formal systems. Since Area II is just that area of intelligent behavior in which digital computers necessarily have the least success, this attempt to treat complex-formal systems as non-formal systems is doomed to failure.

CONCLUSION

What, then, should be the direction of work in artificial intelligence? Progress can evidently be expected in Area III. As Wang points out, we have been given a race of "persistent, plodding slaves" [39:93]; we can make good use of them in the field of simple-formal systems. This does not mean that work in Areas II and IV is wasted. The protocols collected by Newell, Shaw, and Simon suggest that human beings sometimes operate like digital computers, within the context of more global processes. This is really not surprising, since, as Shannon points out, while "most computers are either digital or analogue, the nervous system seems to have a complex mixture of both representations of data" [10:309]. Since digital machines have symbol-manipulating powers superior to those of humans, they should, so far as possible, take over the digital aspects of human information processing.

Thus, to use computers in Areas II and IV, we must couple their capacity for fast and accurate calculation with the short-cut processing made possible by the fringes of consciousness and ambiguity tolerance. A chess player who could call on a machine to count out alternatives once he had zeroed in on an interesting area or in certain parts of the endgame, would be a formidable opponent. Likewise, in problem solving, once the problem is structured and planned, a machine could take over to work out the details (as in the case of machine shop allocation or investment banking). A mechanical dictionary would be useful in translation. In pattern recognition, machines

are able to recognize certain complex patterns that the natural prominences in our experience force us to exclude. Bar-Hillel, Oettinger, and Pierce have each proposed that work be done on systems which promote a symbiosis between computers and human beings. As Rosenblith put it at a recent symposium, "Man and computer is capable of accomplishing things that neither of them can do alone" [10:313].

Instead of trying to make use of the special capacities of computers, workers in artificial intelligence--blinded by their early success and hypnotized by the assumption that thinking is a continuum--will settle for nothing short of the moon. Feigenbaum and Feldman's anthology opens with the baldest statement of this dubious principle:

> In terms of the continuum of intelligence suggested by Armer, the computer programs we have been able to construct are still at the low end. What is important is that we continue to strike out in the direction of the milestone that represents the capabilities of human intelligence. Is there any reason to suppose that we shall never get there? None whatever. Not a single piece of evidence, no logical argument, no proof or theorem has ever been advanced which demonstrates an insurmountable hurdle along the continuum [8:8].

Armer prudently suggests a boundary, but he is still optimistic:

> It is irrelevant whether or not there may exist some upper bound above which machines cannot go in this continuum. Even if such a boundary exists, there is no evidence that it is located close to the position occupied by today's machines [8:392].

Current difficulties, however, suggest that the areas of intelligent activity are discontinuous and that the boundary is near. To persist in such optimism in the face of recent developments borders on self-delusion.

Alchemists were so successful in distilling quicksilver from what seemed to be dirt, that after several hundred years of fruitless effort to convert lead into gold they still refused to believe that on the chemical level one cannot transmute metals. To avoid the fate of the alchemists, it is time we asked where we stand. Now, before we invest more time and money on the information-processing level, we should ask whether the protocols of human subjects suggest that computer language is appropriate for analyzing human behavior. Is an exhaustive analysis of human intelligent behavior into discrete and determinate operations possible? Is an approximate analysis of human intelligent behavior in such digital terms probable? The answer to both these questions seems to be, "No."

Does this mean that all the work and money put into artificial intelligence has been wasted? Not at all, if, instead of trying to hide our difficulties, we try to understand what they show. The success and subsequent stagnation of cognitive simulation and of artificial intelligence in general, plus the omnipresent problem of pattern recognition and its surprising difficulty, should focus research on the three uniquely human forms of information processing. These forms are significantly irrelevant in those two areas of intelligent activity in which artificial intelligence has had its early success, but they are essential in just those areas of intelligent behavior in which artificial intelligence has experienced consistent failure. We can then

view recent work in artificial intelligence as a crucial experiment disconfirming the associationist assumption that all thinking can be analyzed into discrete, determinate operations--the most important disconfirmation of this Humean hypothesis that has ever been produced. In the same way, striking evidence has been collected that not all information can be conceived of in terms of clear and distinct ideas. This technique of pushing associationism and Cartesianism until they reveal their limits suggest fascinating new areas for basic research, notably the development and programming of machines capable of global and indeterminate forms of information processing.

But if the machines for processing informal information must be, as Shannon suggests, entirely different from present digital computers, what can now be done? Nothing directly toward building machines which will be intelligent. We must think in the short run of cooperation between men and digital computers, and only in the long run of non-digital automata which would exhibit the three forms of information processing essential in dealing with our informal world. Those who feel that some concrete results are better than none, and that we should not abandon work on artificial intelligence until some more flexible device for information processing comes along, cannot be refuted. The long reign of alchemy has shown that any research which has had an early success can always be justified and continued by those who prefer adventure to patience.* When one insists on

---

*Enthusiasts might find it sobering to imagine a fifteenth-century version of Feigenbaum and Feldman's exhortation: "In terms of the continuum of substances

a priori proof of the impossibility of success, it is difficult to show that his research is misguided. Artificial intelligence is uniquely vulnerable along this line; still one can always retort that at least the goal can be approached. If, however, one is willing to accept empirical evidence as to whether an effort has been misdirected, he has only to look at the promises and the results.

An alchemist would surely have considered it rather pessimistic and petty to insist that, since the creation of quicksilver, he had produced many beautifully colored solutions but not a speck of gold; he would probably have considered such a critic extremely unfair. Similarly, the person who is hypnotized by the moon and is inching up those last branches toward the top of the tree would consider it reactionary of someone to shake the tree and yell, "Come down!" But if the alchemist had stopped poring over his retorts and pentagrams and had spent his time looking for the true structure of the problem, if the man had come out of the tree and started working perhaps to discover fire and the wheel, things would have been set moving in a more promising direction. After all, three hundred years later we did get gold from lead (and we have touched the moon), but only after we abandoned work on the alchemic level, and reached the chemical level or the even deeper level of the nucleus.

---

suggested by Paracelsus, the transformations we have been able to perform on baser metals are still at a low level. What is important is that we continue to strike out in the direction of the milestone, the philosopher's stone which can transform any element into any other. Is there any reason to suppose that we will never find it? None whatever. Not a single piece of evidence, no logical argument, no proof or theorem has ever been advanced which demonstrates an insurmountable hurdle along this continuum."

BIBLIOGRAPHY

1. Armer, Paul, "Attitudes Toward Intelligent Machines," in Computers and Thought, Edward A. Feigenbaum and Julian Feldman (eds.), McGraw-Hill Book Company, New York, 1963, pp. 389-405.

2. Ashby, W. Ross, "Review of Feigenbaum's Computers and Thought," (manuscript loaned by author).

3. Bar-Hillel, Yehoshua, "The Present Status of Automatic Translation of Languages," in Advances in Computers, Vol. 1, F. L. Alt (ed.), Academic Press, New York, 1960, pp. 91-163.

4. Bullock, Theodore H., "Evolution of Neurophysiological Mechanisms," in Behavior and Evolution, Anne Roe and George Gaylord Simpson (eds.), Yale University Press, New Haven, Connecticut, 1958, pp. 165-177.

5. Descartes, René, Discourse on Method, L. J. Lafleur (trans.), Library of Liberal Arts, New York, 1951.

6. Feigenbaum, Edward A., "Artificial Intelligence Research," IEEE Trans. on Information Theory, Vol. IT-9, November 1963, pp. 248-260.

7. -----, "The Simulation of Verbal Learning Behavior," in Computers and Thought, Edward A. Feigenbaum and Julian Feldman (eds.), McGraw-Hill Book Company, New York, 1963, pp. 297-309.

8. Feigenbaum, Edward A., and Julian Feldman (eds.), Computers and Thought, McGraw-Hill Book Company, New York, 1963.

9. Gelernter, H., J. R. Hansen, and D. W. Loveland, "Empirical Explorations of the Geometry-Theorem Proving Machine," in Computers and Thought, Edward A. Feigenbaum and Julian Feldman (eds.), McGraw-Hill Book Company, New York, 1963, pp. 153-163.

10. Greenberger, Martin (ed.), Computers and the World of the Future, Massachusetts Institute of Technology Press, Cambridge, Massachusetts, 1962.

11. Gruenberger, Fred, Benchmarks in Artificial Intelligence, The RAND Corporation, P-2586, June 1962.

12. Gurwitsch, Aron, "On the Conceptual Consciousness," in The Modeling of Mind, Kenneth M. Sayre and Frederick J. Crosson (eds.), Notre Dame University Press, South Bend, Indiana, 1963, pp. 199-205.

13. Kochen, M., D. M. MacKay, M. E. Maron, M. Scriven, and L. Uhr, Computers and Comprehension, The RAND Corporation, RM-4065-PR, April 1964.

14. Lindsay, Robert K., "Inferential Memory as the Basis of Machines which Understand Natural Language," in Computers and Thought, Edward A. Feigenbaum and Julian Feldman (eds.), McGraw-Hill Book Company, New York, 1963, pp. 217-236.

15. MacKay, D. M., "Internal Representation of the External World," precirculated draft of paper for Avionics Panel, Athens, 1963.

16. -----, "A Mind's Eye View of the Brain," in Progress in Brain Research, 17: Cybernetics of the Nervous System, (a memorial volume honoring Norbert Wiener), Elsevier Publishing Company, Amsterdam, Holland, 1965.

17. Minsky, Marvin, "Steps toward Artificial Intelligence," Proc. of the IRE, Vol. 49, January 1961, pp. 8-30.

18. Newell, Allen, "The Chess Machine," in The Modeling of Mind, Kenneth M. Sayre and Frederick J. Crosson (eds.), Notre Dame University Press, South Bend, Indiana, 1963, pp. 73-89.

19. -----, Some Problems of Basic Organization in Problem-Solving Programs, The RAND Corporation, RM-3283-PR, December 1962.

20. -----, Learning, Generality and Problem-Solving, The RAND Corporation, RM-3285-1-PR, February 1963.

21. Newell, Allen, J. C. Shaw, and H. A. Simon, The Processes of Creative Thinking, The RAND Corporation, P-1320, September 16, 1958.

22. -----, "Chess-Playing Programs and the Problem of Complexity," in Computers and Thought, Edward A. Feigenbaum and Julian Feldman (eds.), McGraw-Hill Book Company, New York, 1963, pp. 39-70.

23. Newell, Allen, J. C. Shaw, and H. A. Simon, "Empirical Explorations with the Logic Theory Machine: A Case Study in Heuristics," in Computers and Thought, Edward A. Feigenbaum and Julian Feldman (eds.), McGraw-Hill Book Company, New York, 1963, pp. 109-133.

24. Newell, Allen and H. A. Simon, Computer Simulation of Human Thinking, The RAND Corporation, P-2276, April 20, 1961; also published in Science, Vol. 134, December 22, 1961, pp. 2011-2017.

25. -----, Computer Simulation of Human Thinking and Problem Solving, The RAND Corporation, P-2312, May 29, 1961.

26. -----, "GPS, a Program that Simulates Human Thought," in Computers and Thought, Edward A. Feigenbaum and Julian Feldman (eds.), McGraw-Hill Book Company, New York, 1963, pp. 279-293.

27. Oettinger, Anthony G., "The State of the Art of Automatic Language Translation: An Appraisal," in Beitraege zur Sprachkunde und Information Verarbeitung, Vol. 1, Heft 2, Oldenbourg Verlage, Munich, 1963, pp. 17-32.

28. Polyani, Michael, "Experience and the Perception of Pattern," The Modeling of Mind, Kenneth M. Sayre and Frederick J. Crosson (eds.), Notre Dame University Press, South Bend, Indiana, 1963, pp. 207-220.

29. Samuel, A. L., "Some Studies in Machine Learning Using the Game of Checkers," in Computers and Thought, Edward A. Feigenbaum and Julian Feldman (eds.), McGraw-Hill Book Company, New York, 1963, pp. 71-108.

30. See, Richard, "Mechanical Translation and Related Language Research," Science, Vol. 144, No. 3619, May 8, 1964, pp. 621-626.

31. Selfridge, Oliver G., and Ulric Neisser, "Pattern Recognition by Machine," in Computers and Thought, Edward A. Feigenbaum and Julian Feldman (eds.), McGraw-Hill Book Company, New York, 1963, pp. 237-250.

32. Scriven, Michael, "The Compleat Robot: A Prolegomena to Androidology," in Dimensions of Mind, Sidney Hook (ed.), Collier Books, New York, 1961, pp. 113-133.

33. Simon, H. A., Modeling Human Mental Processes, The RAND Corporation, P-2221, February 20, 1961.

34. Simon, H. A., and Allen Newell, "Heurustic Problem Solving: The Next Advance in Operations Research," Operations Research, Vol. 6, January-February 1958, pp. 1-10.

35. -----, "Information Processing in Computer and Man," American Scientist, Vol. 52, September 1964, pp. 281-300.

36. Simon, H. A., and Peter A. Simon, "Trial and Error Search in Solving Difficult Problems: Evidence from the Game of Chess," Behavioral Science, Vol. 7, October 1962, pp. 425-429.

37. Smith, D. E., History of Mathematics, Vol. II, Ginn and Company, New York, 1925.

38. Uhr, Leonard, and Charles Vossler, "A Pattern-Recognition Program that Generates, Evaluates, and Adjusts Its Own Operators," in Computers and Thought, Edward A. Feigenbaum and Julian Feldman (eds.), McGraw-Hill Book Company, New York, 1963, pp. 251-268.

39. Wang, Hao, "Toward Mechanical Mathematics," in The Modeling of Mind, Kenneth M. Sayre and Frederick J. Crosson (eds.), Notre Dame University Press, South Bend, Indiana, 1963, pp. 91-120.

40. Wertheimer, M., Productive Thinking, Harpers, New York, 1945.

41. Wiener, Norbert, "The Brain and the Machine (Summary), in Dimensions of Mind, Sidney Hook (ed.), New York University Press, New York, 1960, pp. 109-112.

42. Wittgenstein, Ludwig, Philosophical Investigations, Basil Blackwell, Oxford, England, 1953.

43. -----, The Blue and Brown Books, Basil Blackwell, Oxford, England, 1960.

SUMMARY

Early successes in programming digital computers to exhibit simple forms of intelligent behavior, coupled with the belief that intelligent activities differ only in their degree of complexity, have led to the conviction that the information processing underlying any cognitive performance can be formulated in a program and thus simulated on a digital computer. Attempts to simulate cognitive processes on computers have, however, run into greater difficulties than anticipated.

An examination of these difficulties reveals that the attempt to analyze intelligent behavior in digital computer language systematically excludes three fundamental human forms of information processing (fringe consciousness, essence/accident discrimination, and ambiguity tolerance). Moreover, there are four distinct types of intelligent activity, only two of which do not presuppose these human forms of information processing and can therefore be programmed. Significant developments in artificial intelligence in the remaining two areas must await computers of an entirely different sort, of which the only existing prototype is the little-understood human brain.

ACKNOWLEDGMENTS

I would like to thank Ingrid Stadler, Samuel Todes, Stuart Dreyfus, and Robert Reinstedt for their interest, encouragement, and helpful suggestions.

CONTENTS

SUMMARY iii

ACKNOWLEDGMENTS v

INTRODUCTION 2

Part I

THE CURRENT STATE OF THE FIELD OF ARTIFICIAL INTELLIGENCE 9

SIGNS OF STAGNATION 9
Game Playing 9
Problem Solving 10
Language Translation 12
Pattern Recognition 14
Comments and Conclusions 16

Part II

THE UNDERLYING SIGNIFICANCE OF CURRENT DIFFICULTIES 18

HUMAN VS. MACHINE INFORMATION PROCESSING 18
Fringe Consciousness Vs. Heuristically Guided Search 18
Essence/Accident Discrimination vs. Trial and Error 24
Ambiguity Tolerance vs. Exhaustive Enumeration 30
Perspicuous Grouping--A Derivative of the Above Three Forms 37
Fringe Consciousness 39
Context-Dependent Ambiguity Reduction .. 42
Perspicuous Grouping 44
Conclusion 45

MISCONCEPTIONS MASKING THE SERIOUSNESS OF CURRENT DIFFICULTIES 46
The Associationist Assumption 48
Empirical Evidence for the Associationist Assumption: Critique of the Scientific Methodology of Cognitive Simulation 50

A Priori Arguments for the Associationist Assumption: Conceptual Confusions Underlying Confidence in Artificial Intelligence 55
CONCLUSION 61

Part III

THE FUTURE OF ARTIFICIAL INTELLIGENCE 65
THREE NON-PROGRAMMABLE FORMS OF INFORMATION .. 66
The Infinity of Facts and the Threat of Infinite Progression 67
The Indeterminacy of Needs and the Threat of Infinite Regress 70
The Reciprocity of Context and the Threat of Circularity 71
AREAS OF INTELLIGENT ACTIVITY CLASSIFIED WITH RESPECT TO THE POSSIBILITY OF ARTIFICIAL INTELLIGENCE IN EACH 75
CONCLUSION 82

BIBLIOGRAPHY 87