

ALCHEMY AND ARTIFICIAL INTELLIGENCE

Hubert L. Dreyfus

December 1965

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.

### 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.