

Sophistication in Computers: A Disagreement*

J. L. KELLY, JR.†, MEMBER, IRE, AND O. G. SELFRIDGE‡

Summary—There is frequently more or less acrimonious discussion about artificial intelligence and intelligent machines and their place in science. Usually the discussion settles down to the reiteration of two points of view. This paper is concerned with the difference between them. Do they merely reflect two emotional or ethical biases, or is there an underlying technical judgment on which they disagree? The authors claim the latter and purport to show what that judgment is.

INTRODUCTION

EVERYONE in the field of sophisticated computer programming recognizes that there are developing two antipodal schools of thought on the basic philosophy of *intelligent machines*. The exact nature of this disagreement is hard to specify, but most scientists and many laymen can be categorized as to *pro* or *con*. As we shall see later *pro* and *con* are not suitably descriptive words; nor are *optimistic* and *pessimistic*, nor *modern* and *old-fashioned*. But if the reader finds himself aware of the existence of these two groups, some such pair of words is needed to identify them. We shall use *left-wing* and *right-wing*, respectively, and temporarily leave their meaning as vague as the same words used in politics. Both sides are represented in the authors.

We wish to dismiss as rapidly as possible the crackpots on both sides. Moreover, some right-wingers (and some left-wingers, for that matter) base their opinions on deep and untestable religious grounds and prejudices. We, the authors, not being handicapped by the widespread biases and prejudices of everybody else, hope to explore the question and to show that underneath the dusty irrelevances and traditional pertinacities there really is a genuine scientific question on which scientists may properly differ. For example, the authors do. In this paper we try to uncover exactly what we differ on.

CAN MACHINES THINK?

One question about which thousands of words have been written (and more spoken) is "Can machines think?" While the responses (not necessarily *replies*) to this question are in fact highly symptomatic of left-wingedness or right-wingedness, the wide-spread disagreement about its meaning suggests that it does not fundamentally underlie the title dichotomy. However, we shall explore this straw man further.

First we hope to clear away some of the overlying hindrances. We shall not deal long with the suggestion that the thinking of man is inherently unduplicatable and unique as a matter of faith. First of all, such a suggestion is unprovable and not undeniable, and hence not very helpful or useful to scientists. (We remark that, even

if it were unprovable and undeniable, that is, untestable, we should not therefore deem it *meaningless*, though some would.) Secondly, and more powerfully, people who maintain such a suggestion run a real risk of being demonstrated wrong. Thirdly, and most powerfully, statements which are merely assertions of faith are irrelevant to scientists doing their business of science.

Again, we do not concern ourselves with the argument that man is a machine because all his component follow well-known or less well-known physical laws. It is worth noticing that the direct evidence for this, especially in his brain, is skimpy, though we do not ourselves doubt it. The point is that it is irrelevant to whether we can simulate man with a probably much smaller and certainly very different machine.

Next we say that our differences are not merely semantic. Everybody seems to be aware that the meanings of words change. Even today, we can find a certain reluctance to call an activity *thinking* if it is evident that it can be done by machines. Furthermore, it is just not true that to be able to answer a question one must necessarily be able to define its terms either accurately or precisely. Any man who answered his wife's "Do you love me?" with "What do you mean by *love*?" would deserve what he got. Biologists are not greatly hampered by being unable to define, or even agree about, the meaning of *species* (or, for that matter, *life*). In fact, such disputes often turn out to be rather productive.

It is worth pointing out that in the question "Can machines think?" all three words are ambiguous enough to cause dissension. The word *can* itself, for example, demands clarification.

There are three degrees of possibility; a task is logically possible if it can be described free of contradiction; physically possible if it would not violate the accepted laws of physics; and practically possible at a particular time if it can actually be accomplished with current techniques and materials. These terms are all somewhat relative, but the last category is especially changeable. Examples:

It is impossible (logically) to find an integer between one and two.

It is impossible (physically) to build an antigravity space ship.

It is impossible (practically) to make good synthetic orange juice.

Now we believe that it is certainly logically and physically possible for a digital computer to do any sort of information processing¹ that a man can. This includes

* Received by the PGIT, October 13, 1961.

† Bell Telephone Labs., Murray Hill, N. J.

‡ Lincoln Lab., Mass. Inst. Tech., Lexington, Mass.

¹ By "information processing" we mean tasks whose performance can be measured by means of conventional I-O gear. We will not ask a computer to dig a ditch or eat a hamburger.

thinking or invention regardless of how broadly they are defined. As to what is practically possible, we will cover that in the following paragraphs.

We conclude therefore, that the question on which we differ is not "can machines think?"; that question we save for less sober and less meaningful occasions.

THE TECHNOLOGICAL TIME-SCALE: THE STATE OF THE ART

Even among those who hold that machines can think, in some sense, there are few, except for a rabid fringe on the left, who hold that they (machines) actually *are* thinking these days. What kind of prognosis can we make about artificial intelligence? In these primeval days, the evidence must be (and is) mostly in the form: there are many factors that comprise intelligence, and no one factor has been shown that machines are now incapable of. Creativity, ability to generalize, general language abilities, learning, have all been cited as abilities that are crucial. It is true that programs have been written and run that exhibit every one of them albeit in a limited domain and with special examples. How far can we extrapolate from those examples?

Thus, it is an important question whether those examples are so small as to be insignificant; but this question does not seem to be the functioning one for the disagreement. Nevertheless, we shall discuss its implications, because they seem to us significant to today's problems and are often misunderstood.

The first underlying implication is that the abilities we have been simulating have anything to do with the general question. While it might seem excruciatingly obvious that generalizing and learning have a great deal to do with powerful sophistication in computers, the fact remains that it is only a guess. It might turn out that the truly important abilities are ones we now discount or even some that we cannot describe and that the experiments the left-wingers think to be so significant are indeed irrelevant.

A second underlying assumption is that the complexity of the intelligent computer program of the future is *not* of the essence. Most competent computer programs of today do not have a very complex structure, and most of us use Occam's razor whenever we can to simplify our tasks. We don't suggest here that this is wrong; but it might just be that extraordinary complexity of organization is the real source of the intelligence and flexibility that we want in the programs of the future.

There is a third much more disturbing assumption, disturbing to all of us for what it lightly suggests about the future of science. Supposing we have all the relevant factors well programmed; supposing we have a truly complex and adequate table of organization; supposing we have a stock of applicable history and experiences to feed our program. Why do we think that we can pull all these things together in a reasonable time and make it work? After all, most of our experience tells us that assembling a complex system and making it run are two

very different matters generally separated by a long period of time. The kind of program envisioned for the future of computers is such a complex system by today's standards that we may well ask whether there is any point in bothering to assemble it at all.

We shall not attempt here to describe fully what we conceive to be the present state of the art and its manifold facets. We shall give one small example, which is historically interesting and goes back more than a hundred years. In the 16th Century, von Kempelen² claimed (fraudulently) to have constructed a chess-playing robot; the remarkable thing is that so many decent thinkers were taken in by a concealed dwarf.

Computing machines can play chess now, but not a respectable game. As a matter of fact, we are not even sure that there exists now a working chess-playing program of any sort.³ There has been at least one, however.

The situation is different with checkers; Arthur Samuel from IBM has a program which plays well and can probably beat most readers. Some (for example, Allen Newell and Marvin Minsky) believe that the 7090, say, will be programmed to play master chess within a space of a few years. Others (for example, John Williams and George Miller) believe not.

It is an interesting and instructive example. There are many times in the course of a game where a player must have recourse to what are essentially routines very like those stored so well in a computer. For instance the classical ending of a king, bishop and knight against a solitary king can easily take more than the allowable fifty moves without practice. Many of the intricacies of the openings and their myriad variations are best memorized; that is not to say that a good working knowledge of the fundamental principles of the openings is not invaluable, because it obviously is—unfamiliar situations will often arise, especially with an opponent who may deliberately strive to produce them. But nearly all master players admit that they have effectively memorized a large number of openings.

Thus it seems likely that a successful (good, master, championship, or what have you) chess program will have several different modes of play. It will certainly need an effective evaluation function to handle totally unfamiliar positions, also an ability to handle trees of plausible moves, and so on. Given the almost negligible experience so far, what is the prognosis?

Here is a curious thing: the dichotomy of left- and right-wing does not seem to match views of anticipated progress. As a matter of fact, the right-wing author thinks

² H. W. Dudley and T. H. Tarnoczy, "The speaking machine of Wolfgang von Kempelen," *J. Acoust. Soc. Am.*, vol. 22, pp. 151-166; March, 1952.

R. Willis, "An Attempt to Analyze the Automatic Chess Player of M. De Kempelen with an Easy Method of Imitating the Movements of that Celebrated Figure," with 10 plates; 1821.

³ It is a curious commentary that programs can be dismantled or lost. It has been said "In the computer art, one never has to program anything twice;" in practice, sad to say, programs *can* break down (from excessive correction, for instance) and *can* be dismantled or lost. We have never heard of a *program* wearing out, though cards and tapes do, of course.

that application of techniques already tested can achieve a successful chess program; hence, it might take no longer than a year or so. The other author, a left-winger, believes that we shall need some new techniques for success (namely, the ones he is strenuously advocating); and he believes that the minimum time is about three years.

Our position may then be compared with pioneers on the edge of a new continent. We have axes with which to cut down trees and build bridges. We have ladders to climb cliffs and boats to cross lakes. We have tested our axes on twigs, and made ladders and boats of paper. Shall we make the opposite shore? In principle we can cut down any tree, but obviously trees several miles in girth will take discouragingly long. We can span any river with bridges or boats in principle, but if the river is an Amazon with a thirty-knot current we may not be able to do it in fact. Then again, and here is a crucial point, the continent may be two light years across.

As pioneers, what we do *not* see is a river of molten lava, which would at one sight make us admit the inapplicability of our tools. Similarly, we do not see any clear-cut or easily statable need for new principles of science. If there is any difficulty of this kind it is more diffuse; we suggest that we need better ways of handling or discussing very complicated systems, for example. It will surely turn out that we discover many new principles; that is to be expected. It is not clear that any one of these new principles will in itself so violently reorganize our approaches that that part of thinking which was impossible before suddenly becomes possible afterwards. We guess, therefore, that it will be by small increments in understanding and abilities that we shall be able to reproduce thinking, and that the machines that do it will bear very little resemblance to the computing machines of today.

THE AREA OF DISAGREEMENT

There is one remaining area in which to seek a definition of the relative position of the two schools of thought—the value to science and technology of the work that has been done on the theory of intelligent machines. Here we strike pay-dirt. The authors disagree with each other on every important phase of this question. One believes that a great amount of generality is possible in the approaches to pattern recognition and similar problems; the other insists that each problem must be attacked separately and that there is no stronger connection among them than the mere fact that the working of one problem serves as a not very profound practice for others. One author believes that although there are many useless papers published on intelligent machines (≈ 90 per cent), there do exist some of value (≈ 10 per cent); the other finds no paper on general theory helps in the solving of any particular problem, and in fact does not recognize anything worthy of being called a theory at all at present. One of us believes the concept of machine *learning* is extremely important; the other, that it is a completely

valueless synonym for certain search and parameter adjustment techniques. One believes strongly in the possibility of a genuine scientific theory of intelligent machines and that present work at least scratches the surface; the other is still undecided on even the possibility, but is convinced that nothing of value has been described yet; that pseudo-theories based on “adaptive behavior,” “advice-taking,” “hill-climbing,” and the like are too vague to serve as an actual recipe for writing intelligent programs and too naive to serve as useful heuristics for the programmer.

Man's ability to generalize has been the subject of wonder and speculation for ages. Leibnitz sought a “logic of invention” as distinct from classical deductive logic. Members of the left believe that with the aid of modern computers they have made progress, while their opponents think their efforts trivially transparent. Stated in its baldest form this problem of generalization is appalling to the mathematician. It is simply this: “Given a finite sequence of integers what is the next?” (We leave it to the reader to verify that any *learning* or *generalization* problem can be cast in this manner, although it need not be!) Consider a person playing this game exactly as stated, *i.e.*, a sequence of, let us say, fifteen integers is given him and he guesses the next one. Now we all know that for certain sequences the response will be consistent for nearly all subjects. Now suppose we tried to write a computer program to imitate the person. An extreme right-winger would throw up his hands and declare we could never match human performance. A sensible right-winger would use his intuition to think of easily programmed rules which would model the subject's behavior, but would not attach much general value to his efforts. A sensible left-winger would attack the problem similarly but would claim his method had general value for other tasks. A member of the extreme left would not attack the problem directly. He would consider the input to be a long string of integers representing the subject's entire experience from birth (or perhaps from the beginning of life on earth) and try to duplicate the entire learning process.

We both realize that there is something inscrutable about man's ability to generalize. Members of the extreme right see no possibility of machines ever emulating this ability. Members of the extreme left think the problem is already solved! They are similar to the naive-realists or mechanists of the 18th century who thought that they did or were about to understand all of nature. More reasonable left-wingers realize that complete understanding of any problem is illusory, but that we can reasonably expect present theoretical investigation to yield deeper insight into the problem of generalization. Reasonable right-wingers on the other hand believe that the first step is yet to be taken toward a general theory and that efforts should be directed toward more specific tasks, and that until someone gets some good ideas he should refrain from cluttering the literature.

