TOWARD INTELLIGENT MACHINES*

J. D. Williams
Research Council
The RAND Corporation

P-2170

December 29, 1960

*Text of a talk given December [illegible], 1960, at the [illegible] National Council of Teachers of [illegible]matics, during its meeting at [illegible], Arizona.

TOWARD INTELLIGENT MACHINES

This happy confluence of interest in mathematics and food brings us to the diet problem. In the formal version, from a list of foods, their qualities, and their prices, one seeks the mix of lowest cost that satisfies the body's needs. However, a colleague of mine wanted to find the mix, regardless of cost, that would satisfy him, while his body tended to vanish. He used a list of 600 foods, introduced the notion of feeling full, and fed it all to a machine. Quickly, the machine proposed he begin the day with 700 gallons of vinegar. If he didn't like vinegar, he could start with 500 bouillon cubes, or with a pound of bran, or with a quart of black-strap molasses.

The story has implications I shall not pursue. I confine myself to the one that machines compute with the speed of light, and with the intelligence of the earthworm. I would like to sketch an impression from life among aficionados who hope to educate the earthworm.

It is difficult to pose problems these days that researchers will admit are impossible. Rather, they provide solutions. For instance, if you want to change the axis of rotation of the earth, just say so. The intelligent machine question is now at the mercy of such tender minds.

Beginning with machines of general intelligence not far from the earthworm, the aim is to climb to the level of the human brain. But it doesn't end there, for if man can match Nature, the prognosis is that he

can then outrun Her.

Let me begin by noting some physical characteristics of the brain. The part that thinks, reflects, and remembers, the cerebral cortex, could be housed in a four-inch box. It holds, say (to make it easy), eight billion nerve cells, so two thousand would lie along an edge of the box. It dissipates energy at the rate of a small light bulb — 10 watts. A cell transacts its affairs by firing an electric impulse at some thousands of neighbors, when stimulated to do so by impulses from some neighbors. The decision to fire depends on the previous history of the cell. The maximum rate of fire is about one hundred times a second.

The fine structure of the cortex appears chaotic. The gross structure has some organization, in that there are regions that receive information from the senses, send commands to the muscles, and regulate the body's chemistry. Otherwise, it does not have specialized regions for specialized thinking. The higher functions, such as memory, are distributed throughout; thus memory of this occasion cannot be removed by removing a specific piece of the cortex.

No one is eager to make a copy of this mechanism; the apparently meaningless maze of connections among dubious-looking elements does not inspire Chinese copying. However, about ten years ago some estimates were made of what would be involved in building a machine with so many elements. The requirement for vacuum tubes could have been met by a heap about the size of the Empire State Building, though the machine would

doubtless have covered much of Manhattan. The electric power of the United States could have driven it. Ten thousand tubes would have burned out in a second; though this might not have been serious for a while, for the machine — like the cortex — could tolerate many neuron failures.

The hardware prospect today is much more inviting. We can picture, seriously, elements by the billion, powered by something like a flashlight battery; elements too small to be seen with an optical microscope; elements that fire a million times, or a hundred million times, a second; elements that last indefinitely. The hooker is that we don't know how to do it. More accurately, we are ignorant of what to do.

Let me say a little about today's computer. This is a brain, though not an exciting intellectual generalist. The cortex has a thousand times as many nerve cells, but those of the computer are ten thousand times more speedy and reliable. The nerve cells here, as in the cortex, transact their affairs by electric impulses, which may be interpreted as symbols in the most general way. Those of the computer are in fact interpreted, usually, as numbers, or as instructions pertinent to arithmetic, the task it is deliberately organized to do. It can do as much arithmetic in a minute as a man in a lifetime. Moreover, a man would make millions of errors during his lifetime, whereas if the machine made one after so little work, we would cut off the rent.

Why is it that the general intellectual level of the machine is moronic? There may be more and better reasons than we know, but its

difficulties seem to stem in part from its virtues. For instance, its nerve cells are too simple and reliable: they almost always respond to a given stimulus, or never respond to it; thus they do not benefit from experience. Again, it is so thoroughly organized to do certain tasks well as to preclude doing others — almost.

But that "almost" is a hole through which the researcher may get one of his best handholds: He can teach one machine to simulate another that does not exist. For instance, he can instruct a machine to behave as if a thousand of its simple neurons were a single neuron in a hypothetical new machine. He pays for this: If the actual machine has a million neurons, the simulated machine will have only a thousand. This is a serious loss, for while he began with a thousand-fold deficit, compared to the cortex, he is now down by a factor of a million. Also, the simulated machine will be slower than the original, because it takes time for the simulated neurons to do their housekeeping. Thus the simulated machine will be much smaller, much slower, and — if it is a success — smarter than the original. But now one may build a real new machine, whose basic organization is that simulated, awkwardly. One may build-in the smartness, and also recover the size and speed; and then one can repeat the process. This is the wild, happy world of positive feedback, of geometric series, of compound interest; where the higher one climbs, the better climber he becomes.

You may liken the problem of attaining intelligent machines to that of climbing a mountain; a grotesque, surrealistic mountain protected by

taboos. Present research is like exploratory assaults on the lower flanks, by climbers who do not agree as to the best approach; it is probably a good thing they don't, as yet.

Part of the modest effort is the group of skeptical sidewalk superintendents. This part is perhaps more useful than the climbers suppose, for it discourages the fainthearted, and goads the dedicated and the mad. It is also a promising source of recruits. The actual climbers include natives, such as anatomists, who pry about to discover how the brain is constructed; physiologists, who study the processes of the _living_ organism; and those psychologists who seek to understand the phenomena of behavior. There are mountain-climbing buffs from abroad, such as physicists, chemists, engineers, mathematicians, and philosophers. There are the usual sources of friction between natives and foreigners, and among foreigners: notably language and religion.

It may be of incidental interest to mathematics teachers that the motley crews of experimental psychologists, engineers, mathematicians, and what-have-you, who study the brain under the unpromising title "heuristic research," are fond of mathematics problems and of games such as chess as research vehicles. These tasks are well-suited to finding — through experiments and introspection — the apparent processes used by the brain to solve problems; the idea is then to teach machines to behave in this manner. This is more sophisticated than teaching a machine to play chess as we have taught it to do arithmetic; for here the aim is to

teach the machine to think, and chess is used as an examination.

Perhaps I should add an aside about chess, since it is often mentioned in this context. A bright, inexperienced player can win against a machine, now — sometimes through gamesmanship; a colleague of mine won a game, that he should have resigned, because he knew the machine had not been taught to play the end game. But educated guessers believe the machine will be world's champion some day; perhaps in ten years. One should understand that machines cannot examine all possible chess games. If a machine, as old as the universe, could examine a game in the time it takes an electron to travel its own diameter at the speed of light, the number of games perused would now have about 40 digits. But the total of all chess games is a number with more than 100 digits, so the machine might not yet have found a single winning sequence. Thus good chess is more than taking the best from blindly selected alternatives.

Precisely how the cortex functions is a topic of speculation. A few years ago, the telephone switchboard with its explicit circuitry seemed a useful analogy. The most respectable present guess leans toward statistical mechanics; toward an aggregate of cells through which sweep interacting waves of activity — perhaps like the surface of a pool where the trout are feeding; that thoughts, such as the thought cat, are represented by statistical characteristics of the aggregate, rather than by the states of any specified neurons. Indeed, that if the thought is repeated, the neurons do not necessarily all return to the states they were in the first time. This

conception of the cortex is remarkable, among theories in this field, in that it does not violate the facts, so far as they are known. It is even fruitful, in that it has led to verifiable predictions, such as the complicated wave pattern I have likened to a pool with trout. It makes generous provision for memory: von Neumann estimated that one may receive from the senses, during a lifetime, something like 10^{20} bits of information. Using the pool–and–trout analogy, one could accommodate this quantity by having 20 trout and 20 regions, each region capable of 10 distinguishable states; or with 70 trout, 70 regions, and two states.

It may seem that we are overly concerned with understanding — well enough to imitate it — how man thinks. If the aim were just to have a rough duplicate of the brain, it might be easier to have a baby, so there must be other motivations. Of course the personal motivation of most of the climbers is the classical one: they enjoy climbing. A more objective motivation is that we admire some of the things the brain can do, and wish our artifacts had similar accomplishments. For instance, it has tricks of storing information, and of retrieving information, which seem phenomenal. Its tricks may be less phenomenal than some believe; for instance, I have difficulty accepting the notion that the cortex retains all the information brought to it by the senses; but, as I said, there are differences of religion, which include differences as to what constitutes evidence. But the tricks are phenomenal enough: For instance, consider that, despite a glut of information behind slow–firing neurons, each of us can find in a second or

two some things we may not have looked for in decades: say, the date Columbus discovered America. Everyone have it? Try that on your 7090. Our communications systems, libraries, and computing machines could evidently profit by an infusion of something from this device.

The cortex is also capable of abstractions as varied as the theory of relativity and the Song of Solomon. It can think about itself, and assign value judgments to what it thinks, such as that its thinking is wonderful. However, not every cortex is wonderful, and the theory of relativity and the Song of Solomon are not exactly representative samples of cortical achievement. I have heard that there are people who, given the question "If you had two cows and lost one, how many would you have left?", cannot deal with it — not because the arithmetic is too difficult, but because they don't have two cows; they cannot entertain a hypothesis contrary to fact. If one could build a machine with power of abstract thought comparable to a typical human being — a mathematician, say — well, a typical mathematician, then — it would be worth doing. Of course artificial mathematicians might be a million times faster than the natural ones, in which case the total production of theorems, and other such goodies, staggers the imagination.

The really interesting possibilities begin to emerge when we free ourselves of the notion of matching the human cortex. Rather, let us take as the goal machines that are designed and built really to think.

There are several signs that suggest — at least to my cortex — that that goal is not too ambitious; though it is ambitious enough.

I have alluded to the hardware developments that are coming to hand — thin films and cryogenic devices, for instance — and to the million-fold superiority of the computer over man, achieved with almost the first designs and the first crude electronic devices. I assume therefore that it will be feasible to build almost anything that we have the sense to conceive.

I am also impressed by the implications of comparative anatomy: The brains of vertebrates differ most conspicuously in degree rather than in kind. To be sure, the mammals are better endowed than the reptiles, but, by and large, there is a sequence of brains which, through runs of many animals, is of gently increasing complexity. As a consequence, there is marked similarity between the elements and structures — and evidently the purposes these serve — in rather dissimilar animals; thus if you were to look through the histologist's microscope, you would have to be a bit experienced to know whether you were looking at the cortex of a cat or of a man. It seems inescapable to me that the brain of man, like that of other vertebrates, is an item, of random design, to meet one basic purpose: survival. There is some doubt these days whether it will in fact meet this criterion, but there is no reason to suppose that it is well designed to perform intellectual man's proudest function: namely, to think. The fact that it has outthought things like saber-toothed tigers is no

evidence that it is particularly apt for abstract thinking. It may be argued that I dismiss the survival-tested mechanism too quickly, and perhaps I do. The question arises when we ponder how to motivate advanced machines. I am confident we will discover means to do so, and I will be astonished if we are reduced to threatening them with tigers.

So now we have several reasons for optimism regarding the final outcome of the quest for intelligent machines: Our hardware capabilities are substantial, and increasing swiftly; we have as a tool the machine that can help us design a more sophisticated machine; the competition, the human cortex, probably was designed for another task; and we have a grasp of many things — such as statistical mechanics and experimental design — that are very pertinent to deliberate design.

We may experience here the onslaught of a new disease; indeed, there are already traces of its symptoms in the literature. Consider that, heretofore, man has reigned alone as an arithmetician. Now the machine reigns — and alone, for man's achievement is too trivial to count. As the phenomenon spreads, how do we live with it? We can do so by defining intelligence, or thinking, so as to exclude the machine achievement. This is an undignified effort, at best; as it would be to say that airplanes can't fly because they can't land, gracefully, in trees. It may lead eventually to a proof that <u>we</u> can't think. Perhaps we should just rest on our dignity, as a fish or a protein molecule does in the presence of man; this course has the merit of a boundless future. But we

can surely do better: namely, we can form a feedback loop with the machine, and, to some extent, go with it. The science of genetics is young — scarcely 50 years old — and is moving swiftly. We may learn how to design our children on the machine.

I believe that man can develop extremely intelligent machines; and that he probably will, if he does not have some misadventure otherwise. This development, and his participation in it, will be his greatest adventure.

Let me end this on a lighter note; a story that has the flavor of a Charles Addams cartoon. The story concerns the latest machine, of the oracle class, which of course knows the answers to all questions. The machine is asked, "Is there a God?" It replies, "There is <u>now</u>."