The Mind of the Machine

On the wall of my office is a world map, computer-plotted and therefore not as beautiful as a draftsman would manage. On it are bold outlines, in eight or ten thousand dots, of the huge plates that make up the crust of the earth, which, when they spread apart or touch together or ride one over the other, generate most, perhaps nearly all, substantial earthquakes. The map embodies that realization, for its dotted outlines of "androides" earthquakes. The map embodies that since their own recordings of earthquakes their own world-wide network of detectors recorded from 1961 to 1967. That, however, was as many as all the earthquakes recorded up to that time. They lost a factor of two, which is not much statistically; they gained the advantage of not having to read and interpret all those obscure German journals.

This is a parable for the computer. Like all parables, it has an internal tension: it gives something to the enemies and to the friends of the computer alike. For the friends it is patent that this superb collection of epicenters delineating tectonic plates is probably the single greatest accomplishment of such synoptic study. For an outsider, it is fascinating to see the outline of the rifts and joints. At last we understand something of the earth in the large. At the same time, so cavalier a dismissal of what shall we say if machines come to think? the way, without which I am sure the Coast and Geodetic Survey and its friends would not have been able to produce so beautiful a map.

Perhaps that is a story we are facing in all disciplines. It is quite possible that a similar scheme with its similar swing of a scythe may eliminate even human theorem secretaries, will produce ten million machine-made mathematical theorems per year. But that should not mean that mathematicians are rendered void. We find in man's history previous episodes of the inanimate simulating the animate, from which man has emerged unscathed.

Life as Motion

"For ancient and medieval people there was not much difference between jugglers, alchemists, mechanicians, leeches, star clerks and all such dealers in magic and gramarye."—Joseph Needham.

We accept the idea of the juggler as half-magician. Yet all he does is exploit the fast reaction time of human beings, and the free-fall dynamics of small objects—plates or cones or bowls—in that striking way which so delights the eye, and has uninterrupted done so for at least 3,000 years, by direct archaeological evidence.

With each age of seemingly magical machines that simulate aspects of life, the question of why life is different from non-life returns on a new level. The ancients' simulations of purposeful motion were magic, until man came to understand that life meant more than the ability to move. In an age of computers, what shall we say if machines come to think?

The hidden mechanism of the Draftsman, created by Henry-Louis Jaquet-Droz and Jean-Frederic Leschot in 1774. An upper clockworks drives a shaft upon which are mounted a set ofcams. The shaft rotates once, then releases a lower clockworks, which pushes the cams upward and then sets the upper clockworks in motion again. Thus the "instructions" encoded upon the cams in a system of polar coordinates can be transmitted in turn to the Draftsman's hand by means of a mechanism in the elbow, which also moves the eyes and head. During pauses in the execution of drawings while the set of cams is being incrementally moved, the Draftsman blows on his drawing to free it of dust.

Two of the Draftsman's repertoire of four drawings are shown on the next two pages—left, a portrait of Louis XV; and right, Mon Toutou, my "bow-wow."

The "androides" of Jaquet-Droz are in the Musee d'Art et d'Histoire in Neuchatel, Switzerland, where they are demonstrated on the first Sunday of each month. The illustrations of two Jaquet-Droz automata—the Draftsman and the Writer—that accompany this article are from the Musee d'Histoire et Its Conservateur, Jean-Pierre Jelmini.

The lesson is quite plain: nobody, not the most single-minded proponent of computer data processing, would say that it all began in 1961, even if our modern compatible data began then. The past was an indispensable prologue; it saw the formation of concepts, the development of techniques, the introduction of instruments, the idea of systematic recording, and so on. All this showed the way, without which I am sure the Coast and Geodetic Survey and its friends would not have been able to produce so beautiful a map.

Technology Review, January, 1973 13
evidence. "Alchemists," of course, are people who change matter from one form to another. "Leeches?" I leave the American Medical Association to deal with them. "Star clerks?" Plainly, dealing with the motions of the heavens was beyond the common work of man, beyond primary production, foreign to the social arrangements which make society. Outside us are the stars, with their extraordinary rhythmic motions; those few persons privy to those unique motions had a special role in society. But "mechanicians" hits hard in a mechanical age.

Is this not striking? In the commonplace, banal arguments of today, we contrast on the one hand the spiritual, the aesthetic, the insightful, and on the other hand the dully mechanical. The mechanician embodies the lowest form of intellectual expression!

Though Needham is a modern man, he is able to put himself into the framework of the mind of an older time. He can view mechanicians as being among those persons who deal with magical things. At a time when there was no mechanical prime mover, before even wind- and water-mills, all motion was animate in origin, save the ineffabilities of the waves, the waterfalls, the clouds. At that time—only 3,000 years ago—mechanical contrivances were not economic devices, common in everyday life; they were neither conveniences nor even time-keepers. They were magic.

It will turn out that as far back as we can go in the archaeological record, we will find mechanicians, quasi-magicians, seeking to simulate in their magical devices two great classes of natural motion: the motion of life and the motion of the cosmos.

The first class, the simulation of life of immediate importance to our topic, is the mechanism that could move by itself. Even today, the feeling of a child who is somehow kept away from a surfeit of machine contact (as was every person of 2,500 years ago) is that motion—on any scale save that of the great semi-cosmic motions of the winds and the waves, or of felled trees and the like—necessarily means life; motion means "animation"; the "quick" as opposed to the "dead." Animals move; birds, fish, insects move; but nothing else on our scale moves of itself. The first type of device, then, is the self-moving artifact, motion created by mechanicians as a philosophical simulacrum of life.

There is a long tradition of found artifacts which represent this. The oldest one I know of is a so-called "Jackal of Anubis" from early dynastic Egypt, a figure of a jackal with a jaw that can be moved from behind by strings. One imagines the image apostrophized in the right way in the temple. Its jaw moves while the priest speaks behind it, giving some special dramatic authenticity to its utterance.

Richard Gregory has made a strong case, based on archaeological evidence and plausibility (we are told that is the historian's method), that in the Egypt of the Middle Kingdom, the faces of some temple statues were made concave instead of convex. In consequence, as one walked by this image hollowed out like the inside of a mask, instead of protruding as does the outside of a mask, one's failure to immediately understand the perceptual data meant that parallactic shifts in viewpoint caused the features to appear to move. It is a striking phenomenon. A similar illusion is seen nowadays in looking at pictures of the moon, attempting to decide whether visual prominences are craters or
bumps. This illusion in the form of a statue perhaps makes the statue appear to possess a strange self-motion: it becomes animated, seemingly alive.

We have a very limited literary legacy of the Greek tradition of mechanism. Everyone knows the famous name of Heron (or Hero), writing near 100 A.D., and his teachers to whom he specifically refers and in whose tradition he was plainly working, who stretch back in time about to 250 B.C. One fragment of Hero is perhaps the best known; it is the only one with an English translation, though the Automata by Hero has a famous version in German and there are fine scholarly studies on all his existent works. Everyone has looked at this most famous figure from the work of Hero: the little steam proto-turbine, a little ball of copper with nozzles coming out, all mounted on an axis. Water inside is heated, and steam jets out so that it spins. What one realizes when one looks at the Automata is that the entire volume is a discussion of what can only be called mechanical magic: theatrical effects made by mechanical movement—statues that appear to drink, brass horses that drink even after their heads are cut off, much more.

It is dangerous to quickly dismiss this as merely an eccentric and bizarre foreign body lodged in Greek thought. Admittedly, it is not what we were brought up to believe. We believe in the purity of the Greek intellectual life, in its freedom from crude mechanism, its grave concern for the high issues of beauty and truth. But scholars have found mechanism even in Plato. In the cosmological section of the Timaeus, Plato describes a model in front of him, a real physical object—a kind of armillary sphere. I cannot say this is a majority view of scholars, but allowing for the extraordinary bias towards the written word which historians must necessarily hold (especially those who translate for us from the Greek), I find it an extremely likely theory. The whole story (in the Republic) of the shadows in the cave, one of the most famous images in all of Plato’s arsenal of parables, can be very easily seen as a slightly displaced description of some actual puppet theater, or shadow theater, such as we find used in folk drama in many parts of the world to this day, not least in Greece.

Each of the only two or three of Hero’s books that we have is a sharp account of a series of theatrical effects, including the mechanism—weight- or water-driven—by which they can be executed. The weights do not fall freely but against a cord which very slowly untwists, allowing a valuable change in tempo.

There is a play in five scenes for the automated theater which we know was performed automatically—that is to say, by automata—for over 400 years in Hellenistic times. The best drawing we have shows a theater of dollhouse size on a pedestal. We will rely here upon Richard Brumbaugh, and upon Hero’s account of the plot.

The curtains open. There appear twelve figures, arranged in three rows: the Danaids, who are repairing a ship and moving it forward to be launched into the sea. These figures move busily; one is sawing, others are hammering, yet others work with large and small hole-boring tools. There is a great noise, as of the sound of actual working. After a predetermined time the curtains close; then, opening again, another scene is revealed. One sees the ship of the Achaeans launched in the sea. After the curtains have closed and opened once more, there is only empty sky and painted sea in the background. After a short time, ships sail into view in a line. As one disappears, another enters; dolphins swim along with them, diving into the sea, then becoming visible, just as they really do. There are three dolphins attached to three arms, which pivot just below stage level, so that as a shaft turns, the dolphins leap out of the sea, and dive in again. A very nice effect! Shortly thereafter the sea becomes stormy and the ships run with sail close hauled. When the curtains have once more closed and opened, nothing of the ships can be seen, but there stands Nauplius the King with Athena, and fire burns above the stage. The scene again changes. The shipwreck of Ajax’s boat appears, and Ajax swimming. A machine raises Athena out of view; thunder crashes and a bolt of lightning falls directly from above onto Ajax, who is made to disappear.

In falling, the lightning bolt twists up a cord, so that as soon as it is down the untwisting will bring it back up again sharply. So, zip, zip it goes, down and up again, and in
Though it is more complicated than the Draftsman, the Jaquet-Droz Writer was made two years earlier, in 1772. It contains three sets of 40 cams each, all mounted on the same shaft. A combination of pressures exerted by three cams—one from each of the sets—on three levers guides the Writer's hand in forming letters, and varies downward pressure sufficiently to create light and heavy strokes. A two-thirds revolution of the cams produces a letter, and the remaining third brings a stud of varying height into contact with a mechanism that spaces letters properly.

While this happens, a disc operated by a second clockworks is rotating. Around the periphery of this wheel are 40 wedges, each of which in turn comes into contact with the shaft, pushing it to a height required to bring the three cams required for a chosen letter into contact with the levers. By setting the heights of the 40 wedges, any text can be chosen.

Other mechanisms in the Writer end its operation after it punctuates its sentence with a period, and induce the arm to dip the pen into an inkwell and then shake the pen twice.

At right, the Writer's calligraphy.

Clockwork Life

A simulation of animate life and a mirroring of cosmology were fused in the time of the Enlightenment, when clockwork, stimulated by the enormous development of precision craftsmanship and understanding in Europe, succeeded in producing for the first time mechanisms capable of fulfilling this program of the ancients in full. The near-absurd example is the cuckoo clock, but I hope that we will instead recall the wonderful cam-driven clockwork analog machines of the late eighteenth-century Swiss automaton makers whose expertise forms the ancestry of the watchmakers of today in Neuchatel, Switzerland.

There, on the first Sunday of every month, a curator proudly displays the automata of Jaquet-Droz: figures, two-thirds life-size, beautifully dressed in the costumes of the time, driven by quite elaborate cam-work. One figure is capable of writing, in a fine, handsome, proper little French Alexandrine verse authored by Droz, evidently a talented fellow. Another figure, a young girl, plays on the harpsichord a little piece also composed on the side by Droz. A draftsman draws line figures of Cupids hauling chariots, profiles of Louis XV—four pictures in all.

These were the last of a tradition. Forty years before, de Vaucanson had made a real hit in the early Enlightenment courts, especially at the court of the Sun King himself, with a series of rather droll, not to say raffish, automata. The most famous of these was a duck which was presented to Louis and his courtiers. It quacked, flapped its wings, bent over, ate grain, and then excreted, all there on the table. Let the biochemists laugh a little bit; on a gross level, life was simulated.

As recently as the nineteenth century and in the relatively mundane technology of mechanisms devised that brief time the poor Ajax figure is pulled away. Thus does the clockwork story of the Gods' vengeance upon a murderer come to a conclusion, as stage doors close!
not as wonders but as satisfactions of practical needs, we yet find the fascination with motion of the inanimate. In the nineteenth century, the quality of the new non-wind-propelled, non-horse-drawn, inanimate vehicles was something that still appealed to the popular imagination. I have read that even in the early twentieth century on the plains of Russia and Poland, as you drove by in the very rare automobile, the peasantry might come to demand to look under the hood for the treadmill-pacing animal that they fully expected to find inside.

But for us the magic has gone from clockwork. It is now only some kind of humdrum putting together of gears. It has in principle all been solved, and that’s it. So the fashions go with the centuries. But there was once a time when gear trains seemed a magical simulation of life—just as the phenomena might be that we anticipate or fear coming out of the powerful digital computer of the future.

The Simulation of Intelligence

So far, prolog. Now, the current issue:

Mankind once distinguished life from non-life by purposeful motion, until seemingly magical machine simulations voided that criterion. Here and now, in an age of computers, how shall we understand the uniqueness of life? And what will we have to say of our own minds if machines come to think?

The eighteenth-century simulacra of life established once and for all that it is not by intricate, purposeful motion that living matter is unique. We see now that nor is life unique in having many internal states, for our electronic machines have that. The uniqueness of life is not in the irreversible separation of parts of the organism either, though 40 years ago, this was a life/non-life distinction: in any machine take a part out, and put it back again, and the machine works. Take a part out of a man, and put it back in again; he does not work. By now this seems a very naive criterion, wrong on both sides, but 40 years ago it was a true distinction.

Introspective self-examination? We do not have much of that in our machines. We do not now have machines that reflect even in a small way upon their inner states. They ought to, and certainly if machine design is done in a philosophical spirit, that is the one property that ought not to be left out. Even now, the mere existence of memory-rich machines makes it extremely hard in reality to ignore inner states and accept any stimulus-response theory, any input-output view in which the processing takes place in a "black box" about which nothing need be known. This is not likely to be adequate to represent a machine, a mammal, or a human mind. So much depends on the memories on a computer’s discs that the black-box approach of the behaviorist never occurs to a machine repairman. The machine is a functional of its history. The repairman would be quite lost even in the presence of a simple machine without some sense of what I would like to call the introspective concerns of that machine—

No, that is not quite true yet; I have jumped the metaphor one level more than I am entitled to for the present. But the gap seems only an economic one. In the efforts ahead in artificial intelligence, I would demand, as would an old psychology test, that some trace of self-examination be provided. A machine lacking that goes far toward overlooking the principal philosophical differences between behavior, and behavior with consciousness.

Machines and Mind

We must now ask the question: “Can a machine be comparable to the human mind?” That is what I have surely been preparing to consider. I think I have given enough background to prove this is a legitimate philosophical question, an important one, moreover one we have already met for five thousand years, on one or another level.

I fully believe that we will find no barrier to success in any aspect of machine simulation, in any feature of full machine reproduction of any canon whatever of human mental states. I think that what looks to be true will probably turn out to be true; namely, that the human mind can be described as a slow-clockrate modified-digital machine, with multiple distinguishable parallel processing, all working in salt water. Yet I will say that all those subtleties philosophers and artists talk about, that make life unique and distinguish it from non-life, are true after all, foolish as they may sound when you are young and enthusiastic, and have a naively positive view of science. I offer a position of tension, a determined yes, machines will simulate life, but...

Mind as Questioner

We begin with a statement from a considerable scientist of our own generation, Sir Fred Hoyle, who very stoutly said to me once that the only important thing in science is to ask a good question. An answer will appear; it is eternally implicit in the process. The person who asks the right question is the truly great scientist. That is exactly what is written on Cantor’s gravestone in Latin, so I can cite even a better authority than Hoyle, namely Cantor himself: the question is the essence of science.

Then the notion that thought can be judged by the answers to an interrogation process is much too low a level of test for respectable mental behavior. I will not be satisfied with the machine that claims to simulate human mental life until it asks important questions, raises new problems. That is what I regard as fully human behavior (or at least part of it): not merely answering, however cleverly, however subtly, however neatly picking up literary allusions to Shakespearean sonnets. Merely...
answering questions is not going to do the job. I can almost conceive of a question-answering machine, but a question-asking machine? That is more of a challenge. On the other hand, such a machine, because of imperfections, because art is long and life is short, will lack some of those complex essential features which together made a whole enduring social being out of ourselves after five billion years of earth's history.

But wait: there are strongly limiting constraints on this oracular delivery!

First, a machine simulating the human mind can have no simple optimization game it wants to play, no single function to maximize in its decision making, because one urge to optimize counts for little until it is surrounded by many conditions. A whole set of vectors must be optimized at once. And under some circumstances, they will conflict, and the machine that simulates life will have the whole problem of the conflicting motive, which we know well in ourselves and in all our literature.

Second, probably less essential, the machine will likely require a multisensory kind of input and output in dealing with the world. It is not utterly essential, because we know a few heroic people—say, Helen Keller—who managed with a very modest cross-sensory connection to nevertheless depict the world in some fashion. It was very difficult, for it is the cross-linking of different senses which counts. Even in astronomy, if something is "seen" by radio and by optics, one begins to know what it is. If you do not "see" it in more than one way, you are not very clear what it in fact is.

Third, people have to be active. I do not think a merely passive machine, which simply reads the program it is given, or hears the input, or receives a memory file, can possibly be enough to simulate the human mind. It must try experiments like those we constantly try in childhood—unthinkingly, but instructed by built-in mechanisms. It must try to arrange the world in different fashions.

Fourth, I do not think it can be individual. It must be social in nature. It must accumulate the work—the languages, if you will—of other machines with wide experience. While human beings might be regarded collectively as general-purpose devices, individually they do not impress me much that way at all. Every day I meet people who know things I could not possibly know and can do things I could not possibly do, not because we are from different species, not because we have different machine natures, but because we have been programmed differently by a variety of experiences as well as by individual genetic legacies. I strongly suspect that this phenomenon will reappear in machines that specialize, and then share experiences with one another. A mathematical theorem of Turing tells us that there is an equivalence in that one machine's talents can be transformed mathematically to another's. This gives us a kind of guarantee of unity in the world, but there is a wide difference between that unity, and a choice among possible domains of activity. I suspect that machines will have that choice, too. The absence of a general-purpose mind in humans reflects the importance of history and of development. Machines, if they are to simulate this behavior—or as I prefer to say, share it—must grow inwardly diversified, and outwardly sociable.

Fifth, it must have a history as a species, an evolution. It cannot be born like Athena, from the head full-blown. It will have an archaeologically and probably a sequential development from its ancestors. This appears possible. Here is one of computer science's slogans, influenced by the early rise of molecular microbiology: A tape, a machine whose instructions are encoded on the tape, and a copying machine. The three describe together a self-replicating structure. This is a liberating slogan; it was meant to solve a problem in logic, and I think it did, for all but the professional logicians. The problem is one of the infinite regress which looms when a machine becomes competent enough to reproduce itself. Must it then be more complicated than itself? Nonsense soon follows. A very long instruction tape and a complex but finite machine that works on those instructions is the solution to the logical problem.

One cannot say that the slogan does justice to the extraordinarily complex structure of the "D.N.A. dogma," in which we have the D.N.A. "tape" and the mechanisms whose instructions are encoded upon the D.N.A. That microbiology is positive science. The rest is slogans, but the slogans are powerful ones. They establish in the language of computer science at least the logical possibility we see realized in life: the self-reproducing system. And they do so by a means that makes evolution possible.

By the time I have described all these attributes of a machine to simulate life, I suggest that I have described something which is more like the human mind than it is like our image of the machine. Yet life is not mocked. Mechanical this simula-
tion of life will nevertheless be—as clockwork motions are mechanical.

Nobody believes today that the slightest insight is offered into the nature of organic life by the fact that a clockwork mouse can run across the floor. But there was a time when the wisest people thought that the existence of a simulation of purposeful motion had to be demonstrated. Now it is only a proposition in the transfer of energy and forces on the floor, and we know too much to accept that as a special property of life.

In the same way, I think, we will see man creating a simulation of other aspects of life, yet without creating a doppelgänger of the human mind.

Aspects of the Simulation

In one respect, the human mind works in a way that hints at the methods of a computer. Abstraction of perceptual information comes very early in the human mind’s processing of sensory data; it begins at the level of the peripheral organs. Images are not transmitted from the eyes to form a picture in the brain so that an homunculus in the cortex can look at the picture to decide what it is. Descartes in his day found that idea problematic. Since Hubel and others, we know that aspects of the image are coded right away—abstract items like diagonal lines. An object running out of the field of view might induce a signal that appears at some point—I don’t know where—in some junction box behind the lateral geniculate, or some other complicated anatomical region. But quite early the inputs are transformed, not into point-for-point spatiotemporal representations, but into a more abstract language, one suitable to the machine.

There are, though, chasms between computer design at present and the apparent workings of the mind. Contextual information concerning the meaning of symbols seems indispensable for any economical machine program that has a chance of being “intelligent.” The notion that it can all be done by logical manipulations of these symbols, without any reference to their contexts, is inadequate, for the presentation of many distinct contexts to the central processing device is characteristic of all living beings, characteristic especially of human beings in a social environment of extraordinary richness. But it is absolutely uncharacteristic of the kind of machines we now design and build. Filling that gap will perhaps be one of the greatest steps.

New Genesis

When machines acquire a diverse, self-knowing, active behavior—which is question-asking—and can evolve—which is based on a society of mechanisms out of which some sort of language grows—no one will need to ask if our salt-water machine works in just the same way. We do not argue about the fact that a little mechanical mouse runs along the floor more easily on wheels, not on legs. It is locomotion, all right, differently realized.

We can have creative and personal machines, structures which will act and reflect, and they will share to a degree the attributes of the kind of persons we ourselves are—attributes which we gained by evolution. If the machine does not share those properties, it probably will fail to attain the special high functions of the mind which I have described. My argument arises only out of a sense of the deep unity of the world, a unity which does not demand similar structure for similar functions, but does demand the kind of coherence to which I offer homage.

I do not myself expect to see, but the world may well witness, four kinds of “life.” First, there is our own kind, life continuous by descent over some four billion years, with a heritage of certain antique biochemical fermentations. There will be a second kind, born in a glass test-tube. It will have a new and discontinuous genesis, on quite a different path of information transfer. Perhaps the amino acids will be the mirror images of our own! There will be life of still another sort on another planet. This may be only the Martian fossil plants of the cold pole, or it might even be a fully-conscious life evolved around another star than ours, made known to us by a marvelous microwave link. Finally, there will be a synthetic device, far from biochemistry. It will not have been designed from the beginning by some human programmer, but begun at a higher logical level by humans, to evolve its subsequent internal hierarchies out of its own structure and experience. Once complete, it will behave in the ways I have outlined, in a manner akin to our own nature. It will not be the same as us. But will it be wholly different? By the strength of analogy and faith in the plenitude of the world do I foretell these beings; not by any surer insight.

The father of those that know, Aristotle, wrote: “Mankind is the measure of all things. The hand is the instrument of instruments, the mind is the form of forms.” He was right, not because man is separate from and above nature, but because human beings are part of nature and have been engendered by nature over several billion years. We can expect then to take from our own behavior lessons which may one day lead us to the synthesis, the wondering synthesis, of machine beings, somehow alike and yet very different from that consequence of a cunning, age-old, half-unerring, yet half-random, chain of evolution.

Suggested Readings

Joseph Needham wrote the quoted remark about the magicians of mechanism in his great Science and Civilization in China (Cambridge University Press, 1954), volume 1, page 197. For the illusory sculpture of the Egyptians’ gods, see Richard Gregory’s The Intelligent Eye (McGraw-Hill, 1970). Greek work and thought in the simulation of life, including the play for the clockwork theatre, is discussed in a little popular book whose lightheartedness and freedom from pretension should not be allowed to conceal its originality and depth: Robert Brumbaugh’s Ancient Greek Gadgets and Machines (Crownell, 1966). The history of automata is contained in the richly figured monograph by Alfred Chapuis and Edmond Droz, Automata (Neuchatel, 1958. Though this English edition is exhausted, a French edition, Les Automates, is available from the publisher, Editions du Griffon, in Neuchatel). Derek Price first put thoughts like those in this article in a paper in Technology and Culture, Vol. 5, No. 9 (1964). He has alluded to these matters in nearly all his works since then.