SCIENTIFIC AMERICAN

Could a Machine Think?

Author(s): Paul M. Churchland and Patricia Smith Churchland

Source: *Scientific American*, Vol. 262, No. 1 (JANUARY 1990), pp. 32-39
Published by: Scientific American, a division of Nature America, Inc.
Stable URL: https://www.jstor.org/stable/24996642
Accessed: 09-05-2020 21:54 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Scientific American, a division of Nature America, Inc. is collaborating with JSTOR to digitize, preserve and extend access to *Scientific American*

This content downloaded from 137.110.40.221 on Sat, 09 May 2020 21:54:21 UTC
All use subject to https://about.jstor.org/terms

Could a Machine Think?

Classical AI is unlikely to yield conscious machines; systems that mimic the brain might

by Paul M. Churchland and Patricia Smith Churchland

Artificial-intelligence research is undergoing a revolution. To explain how and why, and to put John R. Searle's argument in perspective, we first need a flashback.

By the early 1950's the old, vague question, Could a machine think? had been replaced by the more approachable question, Could a machine that manipulated physical symbols according to structure-sensitive rules think? This question was an improvement because formal logic and computational theory had seen major developments in the preceding half-century. Theorists had come to appreciate the enormous power of abstract systems of symbols that undergo rule-governed transformations. If those systems could just be automated, then their abstract computational power, it seemed, would be displayed in a real physical system. This insight spawned a well-defined research program with deep theoretical underpinnings.

Could a machine think? There were many reasons for saying yes. One of the earliest and deepest reasons lay in two important results in computational theory. The first was Church's thesis, which states that every effectively computable function is recursively computable. Effectively computable means that there is a "rote" procedure for determining, in finite time, the output of the function for a given input. Recursively computable means more specifically that there is a finite set of operations that can be applied to a given input, and then applied again and again to the successive results of such applications, to yield the function's output in finite time. The notion of a rote procedure is nonformal and intuitive; thus, Church's thesis does not admit of a formal proof. But it does go to the heart of what it is to compute, and many lines of evidence converge in supporting it.

The second important result was Alan M. Turing's demonstration that any recursively computable function can be computed in finite time by a maximally simple sort of symbol-manipulating machine that has come to be called a universal Turing machine. This machine is guided by a set of recursively applicable rules that are sensitive to the identity, order and arrangement of the elementary symbols it encounters as input.

These two results entail something remarkable, namely that a standard digital computer, given only the right program, a large enough memory and sufficient time, can compute *any* rule-governed input-output function. That is, it can display any systematic pattern of responses to the environment whatsoever.

More specifically, these results imply that a suitably programmed symbol-manipulating machine (hereafter, SM machine) should be able to pass the Turing test for conscious intelligence. The Turing test is a purely behavioral test for conscious intelligence, but it is a very demanding test even so. (Whether it is a fair test will be addressed below, where we shall also encounter a second and quite different "test" for conscious intelligence.) In the original version of the Turing test, the inputs to the SM machine are conversational questions and remarks typed into a console by you or me, and the outputs are typewritten responses from the SM machine. The machine passes this test for conscious intelligence if its responses cannot be discriminated from the typewritten responses of a real, intelligent person. Of course, at present no one knows the function that would produce the output behavior of a conscious person. But the Church and Turing results assure us that, whatever that (presumably effective) function might be, a suitable SM machine could compute it.

This is a significant conclusion, especially since Turing's portrayal of a purely teletyped interaction is an unnecessary restriction. The same conclusion follows even if the SM machine interacts with the world in more complex ways: by direct vision, real speech and so forth. After all, a more complex recursive function is still Turing-computable. The only remaining problem is to identify the undoubtedly complex function that governs the human pattern of response to the environment and then write the program (the set of recursively applicable rules) by which the SM machine will compute it. These goals form the fundamental research program of classical AI.

Initial results were positive. SM machines with clever programs performed a variety of ostensibly cognitive activities. They responded to complex instructions, solved complex arithmetic, algebraic and tactical problems, played checkers and chess, proved theorems and engaged in simple dialogue. Performance continued to improve with the appearance of larger memories and faster machines and with the use of longer and more cunning programs. Classical, or "program-writing," AI was a vigorous and successful research effort from almost every perspective. The occasional denial that an SM machine might eventually think appeared uninformed and ill motivated. The case for a positive answer to our title question was overwhelming.

There were a few puzzles, of course. For one thing, SM machines were admittedly not very brainlike. Even here, however, the classical approach had a convincing answer. First, the physical material of any SM machine has nothing essential to do with what function it computes. That is fixed by its program. Second, the engineering details of any machine's functional architecture are also irrelevant, since different

PAUL M. CHURCHLAND and PATRICIA SMITH CHURCHLAND are professors of philosophy at the University of California at San Diego. Together they have studied the nature of the mind and knowledge for the past two decades. Paul Churchland focuses on the nature of scientific knowledge and its development, while Patricia Churchland focuses on the neurosciences and on how the brain sustains cognition. Paul Churchland's *Matter and Consciousness* is the standard textbook on the philosophy of the mind, and Patricia Churchland's *Neurophilosophy* brings together theories of cognition from both philosophy and biology. Paul Churchland is currently chair of the philosophy department at UCSD, and the two are, respectively, president and past president of the Society for Philosophy and Psychology. Patricia Churchland is also an adjunct professor at the Salk Institute for Biological Studies in San Diego. The Churchlands are also members of the UCSD cognitive science faculty, its Institute for Neural Computation and its Science Studies program.

This content downloaded from 137.110.40.221 on Sat, 09 May 2020 21:54:21 UTC
All use subject to https://about.jstor.org/terms
© 1989 SCIENTIFIC AMERICAN, INC

architectures running quite different programs can still be computing the same input-output function.

Accordingly, AI sought to find the input-output *function* characteristic of intelligence and the most efficient of the many possible programs for computing it. The idiosyncratic way in which the brain computes the function just doesn't matter, it was said. This completes the rationale for classical AI and for a positive answer to our title question.

Could a machine think? There were also some arguments for saying no. Through the 1960's interesting negative arguments were relatively rare. The objection was occasionally made that thinking was a nonphysical process in an immaterial soul. But such dualistic resistance was neither evolutionarily nor explanatorily plausible. It had a negligible impact on AI research.

A quite different line of objection was more successful in gaining the AI community's attention. In 1972 Hubert L. Dreyfus published a book that was highly critical of the parade-case simulations of cognitive activity. He argued for their inadequacy as simulations of genuine cognition, and he pointed to a pattern of failure in these attempts. What they were missing, he suggested, was the vast store of inarticulate background knowledge every person possesses and the commonsense capacity for drawing on relevant aspects of that knowledge as changing circumstance demands. Dreyfus did not deny the possibility that an artificial physical system of some kind might think, but he was highly critical of the idea that this could be achieved solely by symbol manipulation at the hands of recursively applicable rules.

Dreyfus's complaints were broadly perceived within the AI community, and within the discipline of philosophy as well, as shortsighted and unsympathetic, as harping on the inevitable simplifications of a research effort still in its youth. These deficits might be real, but surely they were temporary. Bigger machines and better programs should repair them in due course. Time, it was felt, was on AI's side. Here again the impact on research was negligible.

Time was on Dreyfus's side as well: the rate of cognitive return on increasing speed and memory began to slacken in the late 1970's and early 1980's. The simulation of object recognition in the visual system, for example, proved computationally intensive to an unexpected degree. Realistic results required longer and longer periods of computer time, periods far in excess of what a real visual system requires. This relative slowness of the simulations was darkly curious; signal propagation in a computer is roughly a million times faster than in the brain, and the clock frequency of a computer's central processor is greater than any frequency found in the brain by a similarly dramatic margin. And yet, on realistic problems, the tortoise easily outran the hare.

Furthermore, realistic performance required that the computer program have access to an extremely large knowledge base. Constructing the relevant knowledge base was problem enough, and it was compounded by the problem of how to access just the contextually relevant parts of that knowledge base in real time. As the knowledge base got bigger and better, the access problem got worse. Exhaustive search took too much time, and heuristics for relevance did poorly. Worries of the sort Dreyfus had raised finally began to take hold here

THE CHINESE ROOM

Axiom 1. Computer programs are formal (syntactic).

Axiom 2. Human minds have mental contents (semantics).

Axiom 3. Syntax by itself is neither constitutive of nor sufficient for semantics.

Conclusion 1. Programs are neither constitutive of nor sufficient for minds.

THE LUMINOUS ROOM

Axiom 1. Electricity and magnetism are forces.

Axiom 2. The essential property of light is luminance.

Axiom 3. Forces by themselves are neither constitutive of nor sufficient for luminance.

Conclusion 1. Electricity and magnetism are neither constitutive of nor sufficient for light.

OSCILLATING ELECTROMAGNETIC FORCES constitute light even though a magnet pumped by a person appears to produce no light whatsoever. Similarly, rule-based symbol manipulation might constitute intelligence even though the rule-based system inside John R. Searle's "Chinese room" appears to lack real understanding.

This content downloaded from 137.110.40.221 on Sat, 09 May 2020 21:54:21 UTC
All use subject to https://about.jstor.org/terms
© 1989 SCIENTIFIC AMERICAN, INC

and there even among AI researchers.

At about this time (1980) John Searle authored a new and quite different criticism aimed at the most basic assumption of the classical research program: the idea that the appropriate manipulation of structured symbols by the recursive application of structure-sensitive rules could constitute conscious intelligence.

Searle's argument is based on a thought experiment that displays two crucial features. First, he describes a SM machine that realizes, we are to suppose, an input-output function adequate to sustain a successful Turing test conversation conducted entirely in Chinese. Second, the internal structure of the machine is such that, however it behaves, an observer remains certain that neither the machine nor any part of it understands Chinese. All it contains is a monolingual English speaker following a written set of instructions for manipulating the Chinese symbols that arrive and leave through a mail slot. In short, the system is supposed to pass the Turing test, while the system itself lacks any genuine understanding of Chinese or real Chinese semantic content [see "Is the Brain's Mind a Computer Program?" by John R. Searle, page 26].

The general lesson drawn is that any system that merely manipulates physical symbols in accordance with structure-sensitive rules will be at best a hollow mock-up of real conscious intelligence, because it is impossible to generate "real semantics" merely by cranking away on "empty syntax." Here, we should point out, Searle is imposing a nonbehavioral test for consciousness: the elements of conscious intelligence must possess real semantic content.

One is tempted to complain that Searle's thought experiment is unfair because his Rube Goldberg system will compute with absurd slowness. Searle insists, however, that speed is strictly irrelevant here. A slow thinker should still be a real thinker. Everything essential to the duplication of thought, as per classical AI, is said to be present in the Chinese room.

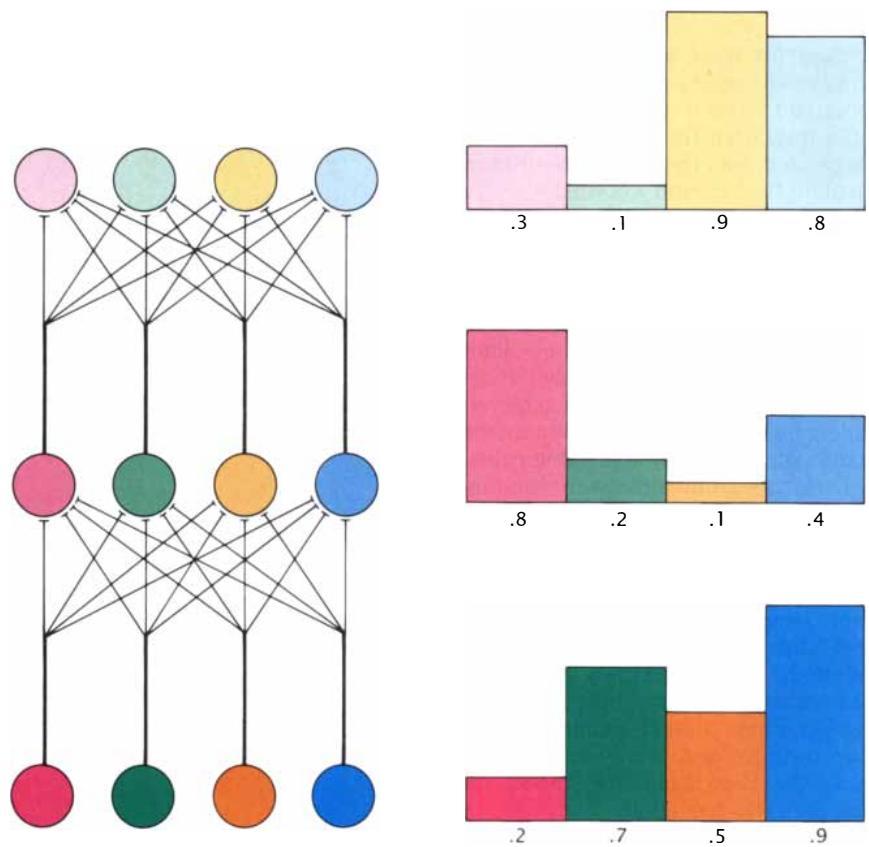
Searle's paper provoked a lively reaction from AI researchers, psychologists and philosophers alike. On the whole, however, he was met with an even more hostile reception than Dreyfus had experienced. In his companion piece in this issue, Searle forthrightly lists a number of these critical responses. We think many of them are reasonable, especially those that "bite the bullet" by insisting that, although it is appallingly slow, the overall system of the room-plus-contents does understand Chinese.

We think those are good responses, but not because we think that the room understands Chinese. We agree with Searle that it does not. Rather they are good responses because they reflect a refusal to accept the crucial third axiom of Searle's argument: "*Syntax by itself is neither constitutive of nor sufficient for semantics.*" Perhaps this axiom is true, but Searle cannot rightly pretend to know that it is. Moreover, to assume its truth is tantamount to begging the question against the research program of classical AI, for that program is predicated on the very interesting assumption that if one can just set in motion an appropriately structured internal dance of syntactic elements, appropriately connected to inputs and outputs, it can produce the same cognitive states and achievements found in human beings.

The question-begging character of Searle's axiom 3 becomes clear when it is compared directly with his conclusion 1: "*Programs are neither constitutive of nor sufficient for minds.*" Plainly, his third axiom is already carrying 90 percent of the weight of this almost identical conclusion. That is why Searle's thought experiment is devoted to shoring up axiom 3 specifically. That is the point of the Chinese room.

Although the story of the Chinese room makes axiom 3 tempting to the unwary, we do not think it succeeds in establishing axiom 3, and we offer a parallel argument below in illustration of its failure. A single transparently fallacious instance of a disputed argument often provides far more insight than a book full of logic chopping.

Searle's style of skepticism has ample precedent in the history of science. The 18th-century Irish bishop George Berkeley found it unintelligible that compression waves in the air, by themselves, could constitute or be sufficient for objective sound. The English poet-artist William Blake and the German poet-naturalist Johann W.



NEURAL NETWORKS model a central feature of the brain's microstructure. In this three-layer net, input neurons (*bottom left*) process a pattern of activations (*bottom right*) and pass it along weighted connections to a hidden layer. Elements in the hidden layer sum their many inputs to produce a new pattern of activations. This is passed to the output layer, which performs a further transformation. Overall the network transforms any input pattern into a corresponding output pattern as dictated by the arrangement and strength of the many connections between neurons.

This content downloaded from 137.110.40.221 on Sat, 09 May 2020 21:54:21 UTC
All use subject to https://about.jstor.org/terms
© 1989 SCIENTIFIC AMERICAN, INC

von Goethe found it inconceivable that small particles by themselves could constitute or be sufficient for the objective phenomenon of light. Even in this century, there have been people who found it beyond imagining that inanimate matter by itself, and however organized, could ever constitute or be sufficient for life. Plainly, what people can or cannot imagine often has nothing to do with what is or is not the case, even where the people involved are highly intelligent.

To see how this lesson applies to Searle's case, consider a deliberately manufactured parallel to his argument and its supporting thought experiment.

Axiom 1. *Electricity and magnetism are forces.*

Axiom 2. *The essential property of light is luminance.*

Axiom 3. *Forces by themselves are neither constitutive of nor sufficient for luminance.*

Conclusion 1. *Electricity and magnetism are neither constitutive of nor sufficient for light.*

Imagine this argument raised shortly after James Clerk Maxwell's 1864 suggestion that light and electromagnetic waves are identical but before the world's full appreciation of the systematic parallels between the properties of light and the properties of electromagnetic waves. This argument could have served as a compelling objection to Maxwell's imaginative hypothesis, especially if it were accompanied by the following commentary in support of axiom 3.

"Consider a dark room containing a man holding a bar magnet or charged object. If the man pumps the magnet up and down, then, according to Maxwell's theory of artificial luminance (AL), it will initiate a spreading circle of electromagnetic waves and will thus be luminous. But as all of us who have toyed with magnets or charged balls well know, their forces (or any other forces for that matter), even when set in motion, produce no luminance at all. It is inconceivable that you might constitute real luminance just by moving forces around!"

How should Maxwell respond to this challenge? He might begin by insisting that the "luminous room" experiment is a misleading display of the phenomenon of luminance because the frequency of oscillation of the magnet is absurdly low, too low by a factor of 10^{15}. This might well elicit the impatient response that frequency has nothing to do with it, that the room with the bobbing magnet already contains everything essential to light, according to Maxwell's own theory.

In response Maxwell might bite the bullet and claim, quite correctly, that the room really is bathed in luminance, albeit a grade or quality too feeble to appreciate. (Given the low frequency with which the man can oscillate the magnet, the wavelength of the electromagnetic waves produced is far too long and their intensity is much too weak for human retinas to respond to them.) But in the climate of understanding here contemplated—the 1860's—this tactic is likely to elicit laughter and hoots of derision. "Luminous room, my foot, Mr. Maxwell. It's pitch-black in there!"

Alas, poor Maxwell has no easy route out of this predicament. All he can do is insist on the following three points. First, axiom 3 of the above argument is false. Indeed, it begs the question despite its intuitive plausibility. Second, the luminous room experiment demonstrates nothing of interest one way or the other about the nature of light. And third, what is needed to settle the problem of light and the possibility of artificial luminance is an ongoing research program to determine whether under the appropriate conditions the behavior of electromagnetic waves does indeed mirror perfectly the behavior of light.

This is also the response that classical AI should give to Searle's argument. Even though Searle's Chinese room may appear to be "semantically dark," he is in no position to insist, on the strength of this appearance, that rule-governed symbol manipulation can never constitute semantic phenomena, especially when people have only an uninformed commonsense understanding of the semantic and cognitive phenomena that need to be explained. Rather than exploit one's understanding of these things, Searle's argument freely exploits one's ignorance of them.

With these criticisms of Searle's argument in place, we return to the question of whether the research program of classical AI has a realistic chance of solving the problem of conscious intelligence and of producing a machine that thinks. We believe that the prospects are poor, but we rest this opinion on reasons very different from Searle's. Our reasons derive from the specific performance failures of the classical research program in AI and from a variety of lessons learned from the biological brain and a new class of computational models inspired by its structure. We have already indicated some of the failures of classical AI regarding tasks that the brain performs swiftly and efficiently. The emerging consensus on these failures is that the functional architecture of classical SM machines is simply the wrong architecture for the very demanding jobs required.

What we need to know is this: How does the brain achieve cognition? Reverse engineering is a common practice in industry. When a new piece of technology comes on the market, competitors find out how it works by taking it apart and divining its structural rationale. In the case of the brain, this strategy presents an unusually stiff challenge, for the brain is the most complicated and sophisticated thing on the planet. Even so, the neurosciences have revealed much about the brain on a wide variety of structural levels. Three anatomic points will provide a basic contrast with the architecture of conventional electronic computers.

First, nervous systems are parallel machines, in the sense that signals are processed in millions of different pathways simultaneously. The retina, for example, presents its complex input to the brain not in chunks of eight, 16 or 32 elements, as in a desktop computer, but rather in the form of almost a million distinct signal elements arriving simultaneously at the target of the optic nerve (the lateral geniculate nucleus), there to be processed collectively, simultaneously and in one fell swoop. Second, the brain's basic processing unit, the neuron, is comparatively simple. Furthermore, its response to incoming signals is analog, not digital, inasmuch as its output spiking frequency varies continuously with its input signals. Third, in the brain, axons projecting from one neuronal population to another are often matched by axons returning from their target population. These descending or recurrent projections allow the brain to modulate the character of its sensory processing. More important still, their existence makes the brain a genuine dynamical system whose continuing behavior is both highly complex and to some degree independent of its peripheral stimuli.

Highly simplified model networks have been useful in suggesting how real neural networks might work and in revealing the computational properties of parallel architectures. For example, consider a three-layer model consisting of neuronlike units fully connected by axonlike connections to the units at the next layer. An input stimulus produces some activation level in a given input unit, which con-

This content downloaded from 137.110.40.221 on Sat, 09 May 2020 21:54:21 UTC
All use subject to https://about.jstor.org/terms
© 1989 SCIENTIFIC AMERICAN, INC

veys a signal of proportional strength along its "axon" to its many "synaptic" connections to the hidden units. The global effect is that a pattern of activations across the set of input units produces a distinct pattern of activations across the set of hidden units.

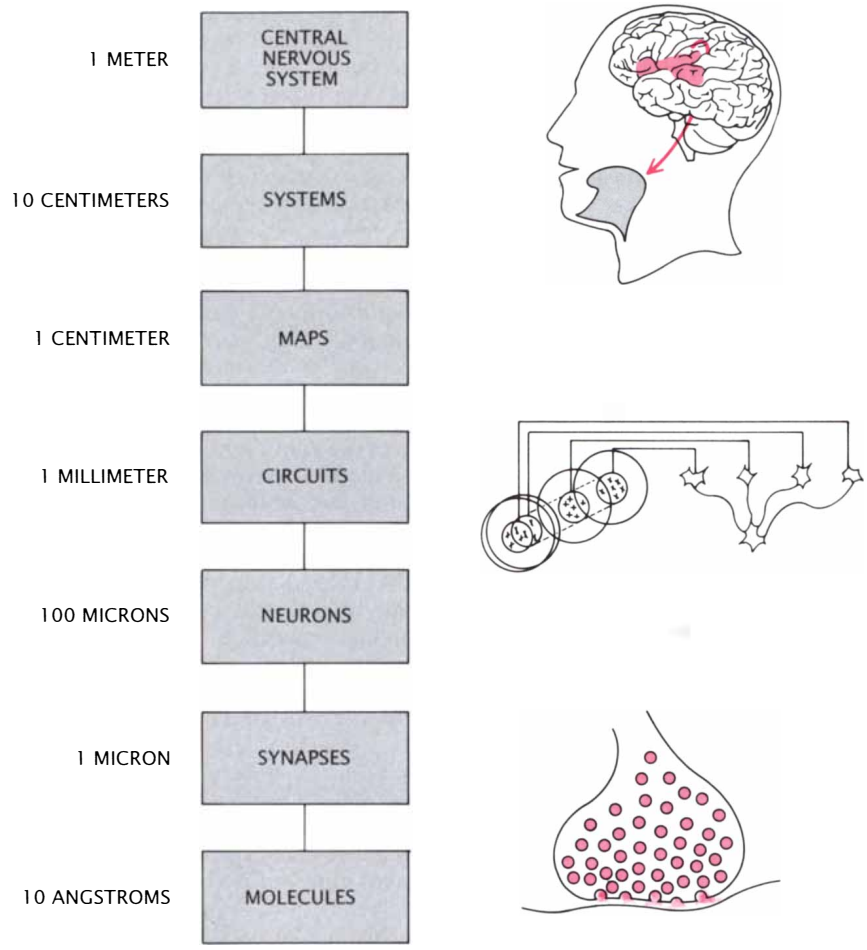
The same story applies to the output units. As before, an activation pattern across the hidden units produces a distinct activation pattern across the output units. All told, this network is a device for transforming any one of a great many possible input vectors (activation patterns) into a uniquely corresponding output vector. It is a device for computing a specific function. Exactly which function it computes is fixed by the global configuration of its synaptic weights.

There are various procedures for adjusting the weights so as to yield a network that computes almost any function—that is, any vector-to-vector transformation—that one might desire. In fact, one can even impose on it a function one is unable to specify, so long as one can supply a set of examples of the desired input-output pairs. This process, called "training up the network," proceeds by successive adjustment of the network's weights until it performs the input-output transformations desired.

Although this model network vastly oversimplifies the structure of the brain, it does illustrate several important ideas. First, a parallel architecture provides a dramatic speed advantage over a conventional computer, for the many synapses at each level perform many small computations simultaneously instead of in laborious sequence. This advantage gets larger as the number of neurons increases at each layer. Strikingly, the speed of processing is entirely independent of both the number of units involved in each layer and the complexity of the function they are computing. Each layer could have four units or a hundred million; its configuration of synaptic weights could be computing simple one-digit sums or second-order differential equations. It would make no difference. The computation time would be exactly the same.

Second, massive parallelism means that the system is fault-tolerant and functionally persistent; the loss of a few connections, even quite a few, has a negligible effect on the character of the overall transformation performed by the surviving network.

Third, a parallel system stores large amounts of information in a distributed fashion, any part of which can be accessed in milliseconds. That information is stored in the specific configuration of synaptic connection strengths, as shaped by past learning. Relevant information is "released" as the input vector passes through—and is transformed by—that configuration of connections.



NERVOUS SYSTEMS span many scales of organization, from neurotransmitter molecules (*bottom*) to the entire brain and spinal cord. Intermediate levels include single neurons and circuits made up of a few neurons, such as those that produce orientation selectivity to a visual stimulus (*middle*), and systems made up of circuits such as those that subserve language (*top right*). Only research can decide how closely an artificial system must mimic the biological one to be capable of intelligence.

Parallel processing is not ideal for all types of computation. On tasks that require only a small input vector, but many millions of swiftly iterated recursive computations, the brain performs very badly, whereas classical SM machines excel. This class of computations is very large and important, so classical machines will always be useful, indeed, vital. There is, however, an equally large class of computations for which the brain's architecture is the superior technology. These are the computations that typically confront living creatures: recognizing a predator's outline in a noisy environment; recalling instantly how to avoid its gaze, flee its approach or fend off its attack; distinguishing food from nonfood and mates from nonmates; navigating through a complex and ever-changing physical/social environment; and so on.

Finally, it is important to note that the parallel system described is not manipulating symbols according to structure-sensitive rules. Rather symbol manipulation appears to be just one of many cognitive skills that a network may or may not learn to display. Rule-governed symbol manipulation is not its basic mode of operation. Searle's argument is directed against rule-governed SM machines; vector transformers of the kind we describe are therefore not threatened by his Chinese room argument even if it were sound, which we have found independent reason to doubt.

Searle is aware of parallel processors but thinks they too will be devoid of real semantic content. To illustrate their inevitable failure, he outlines a

This content downloaded from 137.110.40.221 on Sat, 09 May 2020 21:54:21 UTC
All use subject to https://about.jstor.org/terms
© 1989 SCIENTIFIC AMERICAN, INC

second thought experiment, the Chinese gym, which has a gymnasium full of people organized into a parallel network. From there his argument proceeds as in the Chinese room.

We find this second story far less responsive or compelling than his first. For one, it is irrelevant that no unit in his system understands Chinese, since the same is true of nervous systems: no neuron in my brain understands English, although my whole brain does. For another, Searle neglects to mention that his simulation (using one person per neuron, plus a fleet-footed child for each synaptic connection) will require at least 10^{14} people, since the human brain has 10^{11} neurons, each of which averages over 10^3 connections. His system will require the entire human populations of over 10,000 earths. One gymnasium will not begin to hold a fair simulation.

On the other hand, if such a system were to be assembled on a suitably cosmic scale, with all its pathways faithfully modeled on the human case, we might then have a large, slow, oddly made but still functional brain on our hands. In that case the default assumption is surely that, given proper inputs, it would think, not that it couldn't. There is no guarantee that its activity would constitute real thought, because the vector-processing theory sketched above may not be the correct theory of how brains work. But neither is there any a priori guarantee that it could not be thinking. Searle is once more mistaking the limits on his (or the reader's) current imagination for the limits on objective reality.

The brain is a kind of computer, although most of its properties remain to be discovered. Characterizing the brain as a kind of computer is neither trivial nor frivolous. The brain does compute functions, functions of great complexity, but not in the classical AI fashion. When brains are said to be computers, it should not be implied that they are serial, digital computers, that they are programmed, that they exhibit the distinction between hardware and software or that they must be symbol manipulators or rule followers. Brains are computers in a radically different style.

How the brain manages meaning is still unknown, but it is clear that the problem reaches beyond language use and beyond humans. A small mound of fresh dirt signifies to a person, and also to coyotes, that a gopher is around; an echo with a certain spectral character signifies to a bat the presence of a moth. To develop a theory of meaning, more must be known about how neurons code and transform sensory signals, about the neural basis of memory, learning and emotion and about the interaction of these capacities and the motor system. A neurally grounded theory of meaning may require revision of the very intuitions that now seem so secure and that are so freely exploited in Searle's arguments. Such revisions are common in the history of science.

Could science construct an artificial intelligence by exploiting what is known about the nervous system? We see no principled reason why not. Searle appears to agree, although he qualifies his claim by saying that "any other system capable of causing minds would have to have causal powers (at least) equivalent to those of brains." We close by addressing this claim. We presume that Searle is not claiming that a successful artificial mind must have *all* the causal powers of the brain, such as the power to smell bad when rotting, to harbor slow viruses such as kuru, to stain yellow with horseradish peroxidase and so forth. Requiring perfect parity would be like requiring that an artificial flying device lay eggs.

Presumably he means only to require of an artificial mind all of the causal powers relevant, as he says, to conscious intelligence. But which exactly are they? We are back to quarreling about what is and is not relevant. This is an entirely reasonable place for a disagreement, but it is an empirical matter, to be tried and tested. Because so little is known about what goes into the process of cognition and semantics, it is premature to be very confident about what features are essential. Searle hints at various points that every level, including the biochemical, must be represented in any machine that is a candidate for artificial intelligence. This claim is almost surely too strong. An artificial brain might use something other than biochemicals to achieve the same ends.

This possibility is illustrated by Carver A. Mead's research at the California Institute of Technology. Mead and his colleagues have used analog VLSI techniques to build an artificial retina and an artificial cochlea. (In animals the retina and cochlea are not mere transducers: both systems embody a complex processing network.) These are not mere simulations in a minicomputer of the kind that Searle derides; they are real information-processing units responding in real time to real light, in the case of the artificial retina, and to real sound, in the case of the artificial cochlea. Their circuitry is based on the known anatomy and physiology of the cat retina and the barn owl cochlea, and their output is dramatically similar to the known output of the organs at issue.

These chips do not use any neurochemicals, so neurochemicals are clearly not necessary to achieve the evident results. Of course, the artificial retina cannot be said to see anything, because its output does not have an artificial thalamus or cortex to go to. Whether Mead's program could be sustained to build an entire artificial brain remains to be seen, but there is no evidence now that the absence of biochemicals renders it quixotic.

We, and Searle, reject the Turing test as a sufficient condition for conscious intelligence. At one level our reasons for doing so are similar: we agree that it is also very important how the input-output function is achieved; it is important that the right sorts of things be going on inside the artificial machine. At another level, our reasons are quite different. Searle bases his position on commonsense intuitions about the presence or absence of semantic content. We base ours on the specific behavioral failures of the classical SM machines and on the specific virtues of machines with a more brainlike architecture. These contrasts show that certain computational strategies have vast and decisive advantages over others where typical cognitive tasks are concerned, advantages that are empirically inescapable. Clearly, the brain is making systematic use of these computational advantages. But it need not be the only physical system capable of doing so. Artificial intelligence, in a nonbiological but massively parallel machine, remains a compelling and discernible prospect.

FURTHER READING

COMPUTING MACHINERY AND INTELLIGENCE. Alan M. Turing in *Mind*, Vol. 59, pages 433–460; 1950.

WHAT COMPUTERS CAN'T DO; A CRITIQUE OF ARTIFICIAL REASON. Hubert L. Dreyfus. Harper & Row, 1972.

NEUROPHILOSOPHY: TOWARD A UNIFIED UNDERSTANDING OF THE MIND/BRAIN. Patricia Smith Churchland. The MIT Press, 1986.

FAST THINKING in *The Intentional Stance*. Daniel Clement Dennett. The MIT Press, 1987.

A NEUROCOMPUTATIONAL PERSPECTIVE: THE NATURE OF MIND AND THE STRUCTURE OF SCIENCE. Paul M. Churchland. The MIT Press, in press.

This content downloaded from 137.110.40.221 on Sat, 09 May 2020 21:54:21 UTC
All use subject to https://about.jstor.org/terms
© 1989 SCIENTIFIC AMERICAN, INC

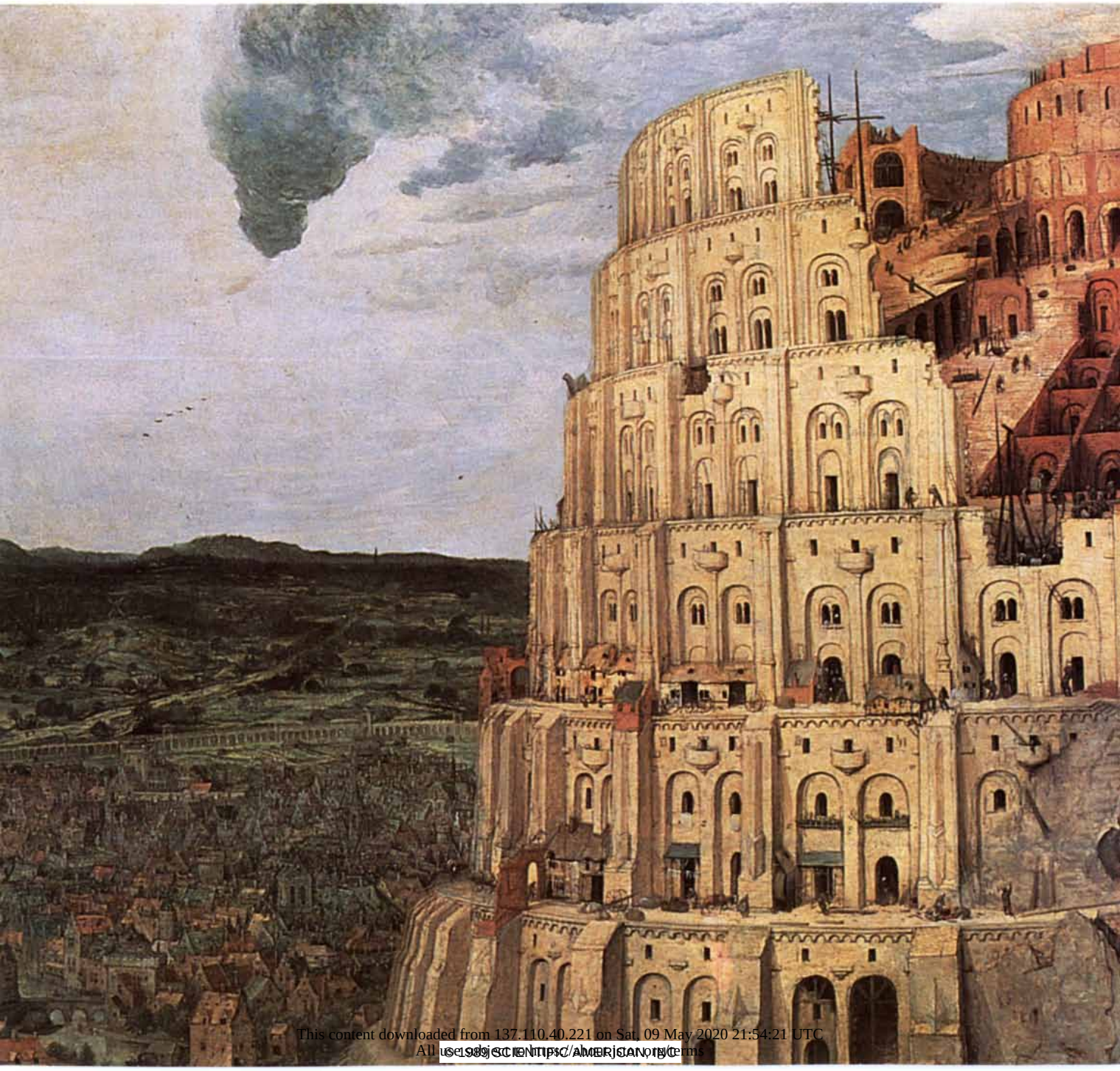
THE TOWER OF BABEL AND SYSTEMS INTEGRATION

The parable is older than writing itself, coming to us from the first murmurings of civilization. Yet its lesson seems to have been aimed specifically at the late twentieth century.

In Genesis, Chapter 11, we read of an unnamed people building a great city on the plain of Shinar (Mesopotamia). To the narrator of this parable, peering across time and desert from his own nomadic traditions, these folk were awesomely clever. They all spoke one common language, and because of this, nothing was impossible to them.

The plan of these ingenious people was to erect a

Pieter Bruegel the Elder, c. 1560, Kunsthistorisches Museum, Vienna



This content downloaded from 137.110.40.221 on Sat, 09 May 2020 21:54:21 UTC
All use subject to https://about.jstor.org/terms

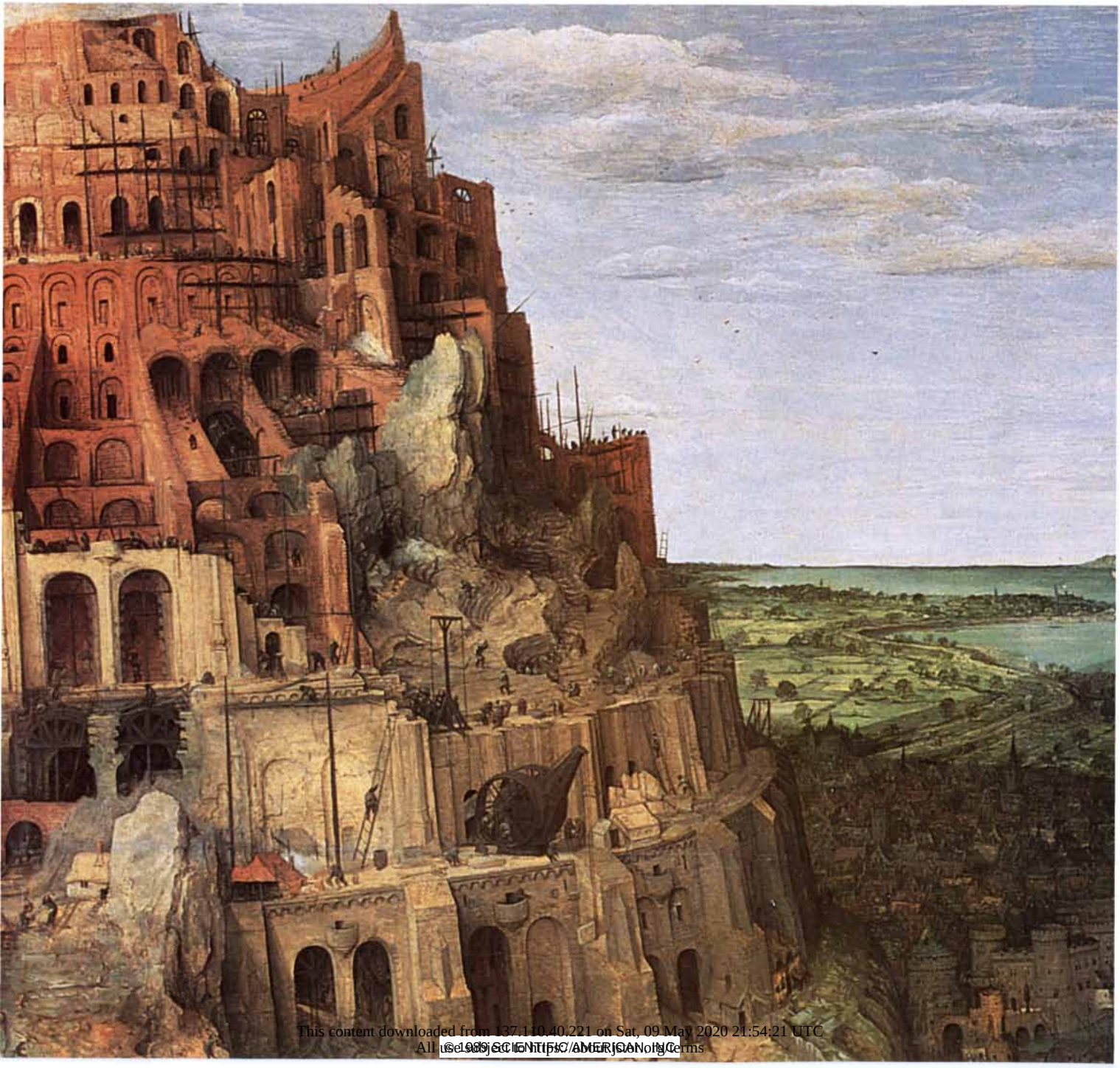
© 1989 SCIENTIFIC AMERICAN, INC

huge temple tower, a ziggurat, whose top would reach into heaven. It was to be an altar to their own intellect and would be called Babel, or "Gate of God." But God himself came down and walked the streets of their city and saw their project under construction. The hubris of this arrogant race angered him. He passed his hand over the city and cursed it. Now where there had been one language were suddenly hundreds. Confusion reigned. Nothing was possible. The people abandoned their city and scattered across the land, taking with them their bewildering tongues. And their vaunted temple, the Tower of Babel, was left untopped; carrion for the wind.

The lesson taught by this ancient parable is uncannily prescient for us in the twentieth century. The revolution in information technology during the past four decades has brought with it the ancient curse of Babel. Every year witnesses the birth of new computer companies, all fiercely competing with faster, more powerful hardware, new formats and new languages. All contributing to an atmosphere of discord that the narrator of the Biblical story would have had no trouble recognizing, despite the great gulf of time.

Recognizing this discord, Lockheed has a solution; systems integration. For years the company has been synthesizing apparently incompatible systems, whether for use in space, the military, or private industry. To this end, Lockheed has actually been able to work against the Babel effect. And with everyone once again speaking the same language, who knows what wonders are possible?

Lockheed
Giving shape to imagination.



This content downloaded from 137.110.40.221 on Sat, 09 May 2020 21:54:21 UTC
All use subject to https://about.jstor.org/terms
© 1989 SCIENTIFIC AMERICAN, INC