24.09x Minds and Machines

Paul M. Churchland and Patricia Smith Churchland, "Could a machine think?"

Excerpts from Paul M. Churchland and Patricia Smith Churchland, "Could a machine think?" (*Scientific American* 262: 32-7, 1990)

The Churchlands start with a question.

> Could a machine think? [By the early 1950s] [t]here were many reasons for saying yes. One of the earliest and deepest reasons lay in two important results in computational theory.

The first result is:

> 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 non-formal 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 result is:

> 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.

The Churchlands first point out a consequence of the two results.

> 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.

They then discuss the program of classical AI in this light.

> 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.
>
> …
>
> 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 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.

Well, perhaps the case wasn't completely overwhelming, because "there were some arguments for saying no", one of which is Searle's Chinese room argument.

> [In 1980] John Searle authored a new…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.
>
> 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 non-behavioral test for consciousness: the elements of conscious intelligence must possess real semantic content.

The Churchlands reject Searle's Axiom 3: *syntax by itself is neither necessary nor sufficient for semantics*. (See Searle, "Is the brain's mind…", p. 4.) They think Axiom 3 is not obviously true; Searle needs to give us reason to think it's true, and he hasn't.

> 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.

Does the Chinese room argument show that Axiom 3 is true? Not according to the Churchlands.

> 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. 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.

They develop what they claim to be a parallel argument with an obviously wrong conclusion.

> 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!"

They consider how Maxwell should respond to this objection.

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!"

Despite the derision, Maxwell is right. And, the Churchlands say, the proponent of Strong AI should give the same response 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.

However, the Churchlands agree with Searle that if the Chinese room is a "rule governed SM machine" it's unlikely to generate any thinking. But that's not because no computer can think, but because the architecture of a classical serial computer is not very brain-like.

> 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. [There have been] 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.

The Churchlands think that machines can think, so long as they are certain kinds of "parallel machines." They explain.

> 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.

They claim that Searle's argument does not target brain-like parallel machines.

> [I]t is important to note that [a parallel system modeled on the nervous systems] 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; [parallel systems] 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.

The Churchlands consider Searle's "Chinese gym" objection.

> 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 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.

Their response:

> 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 [neural network] 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 Churchlands conclude by emphasizing that the brain is a very distinctive kind of computer.

> 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.

They then point out their source of disagreement with Searle.

> 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 brain-like 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 non-biological but massively parallel machine, remains a compelling and discernible prospect.