Remote Learning Module for 15 April 2020

Lecture Notes on *Mind, Matter, and Mathematics* – Chapter 6

Last class took a deeper look into the utility of the Darwinian schema for making sense of how human brains form up mathematical ideas, especially those involved with the discovery of new mathematical objects and truths, with our primary focus trained on neurological correlates to natural selection. Today we'll look to Chapter 6: "Thinking Machines," or the prospect of engineering Artificial Intelligence. Here, we'll consider three flavors of theory: (i) GOFAI (Good, Old Fashioned Artificial Intelligence), (ii) Neurocognitive Science, and (iii) Neuromimetics.

* * *

(1) Are Intelligent Machines Possible?

Let's note at the outset, that both Changeux and Connes are more or less agreed that human brains are intelligent machines, so at least we can say that if the project of Artificial Intelligence was to engineer things exactly like ourselves, well, we already know how to do that. So, our question is not about natural intelligence but mechanical intelligence.

As just mentioned, there are three prospects for accomplishing such engineering short of human biological reproduction: (i) **GOFAI** (Good, Old Fashioned Artificial Intelligence), where we look to *simulate* or mimic the cognitive *functions* of the brain; (ii) **Neurocognitive Science**, where we treat the brain as a thinking machine; and (iii) **Neuromimetics** (or neural networks), where we design machines that simulate brain *structure* (not function), attempting to simulate brain functions.

(2) Gödel's Theorems.

The concern here focuses mostly on the possibility of achieving Artificial Intelligence in the programming of digital computers (Turing Machines). One set of objections to the prospect comes from Gödel's Theorems (the first of which demonstrates the Incompleteness of Arithmetic and the second of which demonstrates that arithmetic contains sentences that can neither be proven true nor proven false from the Peano axioms plus standard first-order predicate logic). On Changeux's and Connes' reading, these theorems provide the grist for arguments of the form: you can't simulate brain functions in a logical system because logical systems "can't define themselves."

The original project for providing secure logical foundations for all of arithmetic was Russell and Whitehead's (in *Principia Mathematica*). Now, as Connes notes, this was not the same project as the one envisaged by David Hilbert, who wanted to reduce arithmetic sentences and proofs to a formal language of symbols and manipulation rules. Russell and Whitehead wanted to show a good deal more: they wanted to show that arithmetic just *was* classical logic—translated into compressed expressions, but in the end reducible to the propositional calculus. It was this project that faltered (fatally) on the Gödel results.

Note well, however, that the rendering of these results on p. 158 of *Mind, Matter, and Mathematics*, is inaccurate: you *cannot*, in fact, arbitrarily assign a truth value to an *undecidable* arithmetic proposition (as opposed to an *independent* axiom). An example of an independent axiom would be the Parallel Postulate in geometry: you can assign the value true to this postulate and you get Euclidean geometry (flat space); but you can deny the postulate without contradiction, whereupon you get hyperbolic or parabolic geometry.

Also, note that Gödel's proofs are presented in a second-order theory: they show that the first-order apparatus of arithmetic generates truths that it cannot itself prove. To get the flavor of the Second Theorem, think of proofs (*Bedeutung* in German) as analogous to beliefs about beliefs. On this analogy, the following sentence must be *undecidable*: "You do not believe this sentence." Why? Because if you do believe it, then you believe it to be true, so you don't; and if you don't then what it says of you is true, so you do. Your belief system must therefore be incomplete: there are truths about your beliefs (for example that "You cannot consistently assert this sentence") that you cannot consistently assert. The best you can do is arbitrarily avoid self-referential usage: there is no consistent logical algorithm for avoiding such sentences though.

(3) Turing's Thinking Machines.

In "Computing Machinery and Intelligence (1950), Alan Turing revisited his 1936 paper (published in January of 1937), "On Computable Numbers with an Application to the Entscheidungs Problem." This problem is otherwise known as the "Halting Problem." The question posed by the Halting Problem is whether or not for any given program or algorithm a suitably programmed computer will halt (deliver an answer) or continue forever without halting. If you ask a program to compute the sum of 2 + 2, it will halt, giving the answer, 4. If you ask instead, for the largest integer, the program is not halt (because there is no largest number). The problem is to be able to say of any program whether or not it will halt. And this question is undecidable.

All the same, Turing proposed in the 1950 paper that if a digital computer was able to answer questions put to it by a human being sufficiently well that the human being could not tell whether s/he was conversing with a machine or an animal, then we might as well say that the machine can think. His point was that the question of machine intelligence should be put to empirical test, rather than being answered apriori by conceptual analysis.

(4) Analogies to the S-Matrix in Quantum Mechanics.

What Heisenberg's new matrix algebra for quantum calculations expresses philosophically is "Hands out of the box!" We *cannot* know the position and momentum of a quantum particle to any arbitrary degree of precision, but we *can* calculate all the quantities we can experientially measure. For Neils Bohr (who is unmentioned by Changeux and Connes), the puzzle presented by this feature of quantum mechanics can be solved by adopting the Principle of Complementarity. On this principle, both position and momentum (classical quantities) are necessary for a complete description of mechanical phenomena, but they cannot be measured in classical apparatus to any arbitrary degree of precision with respect to each other. To take a measurement, we must set up an apparatus, and, therefore, reality cannot be separated from the apparatus that measures it, even though that's how we've got to describe reality (that is, as existing independent of our apparatus). Similarly, thinking cannot be separated from the apparatus that measures it, even though, that's how we want to describe it.

(5) Is the Brain a Computer?

Moving on from GOFAI, Changeux and Connes consider the prospects of neurocognitive science. Here, Connes relies mostly on analogies to playing chess, but his drift falls nicely into place with Damasio's somatic marker hypothesis: the mechanisms involved in higher-order cognition just involve the limbic system, whereby emotions light up the decision tree as we make moves in a game like chess.

(6) A Self-Evaluating Machine that Can Suffer.

The title of this section tells it all in a nutshell. If we want to imitate brain anatomy in hopes of generating brain physiology, then we shall need to acknowledge the differential role of pain in the overall scenarios we expect to confront. Remember Dennett's proviso about the "real predicaments" facing a mind seeking sustenance and homeostasis in a natural environment. There are innumerably more ways to get dead than to stay alive. So, in order to avoid the "slings and arrows of outrageous fortune" that will confront an AI built of neural networks, there will need to be an Evaluation Function analogous to pain responses in animals.

Curiously enough, however, in looking to come up with a local evaluation function from the global one (pain), Changeux and Connes do not take advantage of one of the most carefully laid out maps of the same terrain from the ancient world: Aristotelian *habit*.

* * *

Next time, we'll look to Chapter 7 (The Real and the Rational) as well as the Epilogue (Naturalistic Ethics), as our intrepid interlocutors wrap up their positions and speculate on the relation between mathematics and the moral sciences. Be well everyone, and, although I imagine you are probably quite tired by now of my continuing to say so, do remember: social distancing continues to save lives, which is presumably why we are still not in JUB 202 presently.