Meeting 2: The Halting Problem

CS198: The Poetry of Computer Science, the Computer Science of Poetry
Philosophy of Computation at Berkeley
pocab.org

October 2, 2017

> It is an inherent property of intelligence that it can jump out of that which it is performing, and survey what it has done; it is always looking for and often finding patterns. Now I said that an intelligence can jump out of its task, but that does not mean that it always will. However, a little prompting will often suffice. For example, a human being who is reading a book may grow sleepy. Instead of continuing to read until the book is finished he is just as likely to put the book aside and turn off the light. He has stepped "out of the system" and yet it seems the most natural thing in the world to us. Or, suppose person A is watching television while person B comes in the room, and shows evident displeasure with the situation. Person A may think he understands the problem, and try to remedy it by exiting the present system (that television program), and flipping the channel knob, looking for a better show. Person B may have a more radio concept of what it is to "exit the system" – namely, to turn the television off.
>
> Of course, there are cases where only a rare individual will have the vision to perceive a system which governs many people's lives, a system which has never before even been recognized as a system; then such people often devote their lives to convincing other people that the system really is there and that it ought to be exited from!
>
> (Douglas Hofstadter, *Gödel, Escher, Bach*)

1 The Mathematical Marlboro Man

Today we start delving into formal mathematical concepts, and have to deal with names like "Turing machines", "The Church-Turing thesis", "Gödel's Incompleteness Theorem", et cetera. But before we delve into it, let's examine the system (how math is socially constructed and talked about) in which we're working in, so that we are able to "exit" it if we want to.

> Mathematicians are sometimes portrayed as intellectual cowboys out to tame the mathematical universe – what one might descirbe as a Mathematical Marlboro Man. Indeed, mathematics has been described as "the science which lassos the flying stars." Mathematicians are depicted as living heroic lives, filled with self-sacrifice, all in the name of the search for truth...
>
> Instead of trying to tame horses or cattle, mathematicians tame creatures such as infinity. The mathematician James Pierpont writes, "The notion of infinity is our greatest friend; it is also the greatest enemy of our peace of mind. ... Weirstrass taught us to believe that we had at last thoroughly tamed and domesticated this unruly element. Such however is not the case; it has broken loose again and Hilbert and Brouwer have set out to tame it once more. For how long? We wonder."
>
> (Claudia Henrion, *Women in Mathematics*)

- What do you think of Pierpont's metaphor? What aspects, if any, are true, and what aspects, if any, are false?
- Consider the masculinity of mathematics with respect to the "brogrammer", "lone wolf programmer" culture. In what aspects are they similar and in what aspects are they not?

> In any mathematics journal there may be found language such as that in the following abstract, which bears the title "A Boleslawskian Criterion for the Hughes-Williams Evaluation of non-Walquistness":

Let S be the standard Smith class of normalized univalent Matcuzinski functions on the unit disc, and let B be the subclass of normalized WaIquist functions. We establish a simple criterion for the non-Walquistness of a Matcuzinski function. With this technique it is easy to exhibit, using standard Hughes-Williams methods, a class of non-Walquist polynomials. This answers the Kopfschmerzhaus-type problem, posed by R. J. W. ("Wally") Jones, concerning the smallest degree of a non-Walquist polynomial.

...

[W]hile the place of such words in mathematical discourse is beyond question, what is not beyond question is the widespread practice, as in our introductory example, of recklessly coining and using new eponymous terms, without consideration either to possible alternatives or to likely consequences.

(Henwood & Rival, "Eponymy in Mathematical Nomenclature")

- Write a convincing fake abstract of a mathematical paper.

  TheoryMine lets you name a personalised, newly discovered, mathematical theorems as a novelty gift. Name your very own mathematical theorem, newly discovered by one of the world's most advanced computerised theorem provers (a kind of robot mathematician), and you can immortalise your loved ones, teachers, friends and even yourself and your favourite pets.

  (theorymine.co.uk)

- Suppose you've just started a job as a theorem salesman. Develop your best pitch to sell a theorem.
- Consider the following quote:

  > It takes a thousand men to invent a telegraph, or a steam engine, or a phonograph, or a photograph, or a telephone or any other important thing—and the last man gets the credit and we forget the others. He added his little mite — that is all he did. These object lessons should teach us that ninety-nine parts of all things that proceed from the intellect are plagiarisms, pure and simple; and the lesson ought to make us modest. But nothing can do that.
  >
  > (Mark Twain)

  Some might say this is too harsh: Andrew Wiles, for example, worked by himself for seven years to prove Fermat's Last Theorem. Shouldn't that kind of dedication be rewarded with due credit? In fact, Wiles received more than a million dollars from various prize agencies for his effort. If Twain is right, that money should be distributed among dozens, maybe hundreds, of people. Do you think Wiles deserved that prize or no? In what case does one "deserve" anything?

2 Formal systems and Turing machines

There are many expositions of formal systems. They usually go like this:

- There are a set of *axioms*, which are truths assumed to be true.
- And there are a set of *deduction rules*, by which
- valid *theorems* of the formal system are produced.

It's kind of like a tree: the roots are the axioms, the deduction rules are patterns of how branches grow, and the tip of a branch is a theorem. The entire branch is a proof.

A Turing machine is usually explained in terms of "tapes", "cells", and "transition functions", but really a Turing machine is the exact same thing as a formal system, so once you understand what a formal system is, you've also understood what a Turing machine is:

- *Axioms* correspond to *inputs* of a Turing machine.
- *Deduction rules* correspond to *transition functions* of a Turing machine, by which
- New *configurations* of the Turing machine are created.

It's also kind of like chess. In chess, we agree on the initial configuration of the board. Everyone agrees that the king should be next to the queen, the pawns should be lined up neatly in front row, and so on. We also agree on how a piece can move, and how a piece can capture another piece. When I move a pawn from here to there, the board looks different (obviously). In other words, the board has entered a new *configuration*.

One more metaphor: consider a bunch of colorful balls. Your mom wants you to put the balls in a line, but she is very particular about which ball can come after which ball, and she will be very mad at you if two balls are in a bad order. She gives you a book of rules (axioms and transition functions): a blue ball, but not a brown ball, can come to the right of a red ball; a ball to the right of some ball can't be larger than that ball; the first ball must be soft and squishy; if you placed a black ball, stop placing any more balls; and so on. You faithfully place the balls, and your mom comes to inspect them. Suddenly, she sees a green ball to the right of a red ball, and this configuration of color brings up some forgotten trauma which eminently displeases her. She would like to whip you, but alas, she never wrote in her rule book that a green ball can't come after a red ball! That is, your correct order of balls is a *theorem*, and you have a *proof* that they are indeed in the correct order – just inspect each ball, one after another, and you can show your mom what rule you used from her little rule book to get from one to the next. So you get off free, and you are are happy.

The point is that the transition functions are rules we've agreed on beforehand. And because we need to start *some*where, we also agree on what axioms to use. So clearly, which rules and axioms we agree on beforehand must effect what kinds of conclusions we can reach. As in: if mom *did* have a rule saying that green can't come after red, you'd be in for a whipping. But as it turns out, and it may be difficult to grasp this concept, at a certain point, it doesn't matter which rules we use!

The caveat is, of course, "at a certain point". Suppose your mom really likes this ball arrangement business and would like to have you do it for an infinite amount of time, and she wants you to produce an infinite number of ball configurations. Now she must take care in writing her rule book, because she doesn't want you to ever run out of configurations, and making you create a configuration you've already created would be just *cruel*. So what kind of rules should she devise? She can't have some rule that limits the number of configurations you can make, like: start with a white ball, always put a red ball after a white ball, and upon reaching a red ball, stop placing balls down. This would make only one configuration. That is, her rulebook is not powerful enough. The rulebook she wants should have rules capable of generating an infinite number of ball-configurations. Such rules exist; and, again, it may be counterintuitive, but exactly what the rules are don't matter at this point. In slogan form, the most powerful sets of rules are all the same. Each is exactly as powerful as any other. In a way, the power of the rules emerge when they are taken as a whole, without any one rule mattering much. We call this power by the eponymous phrase, "Turing-complete". So your mom can torture you for infinity, no problem.

3 The Church-Turing Thesis

> The Church-Turing thesis says that any real-world computation can be translated into an equivalent computation involving a Turing machine.
>
> (Wolfram MathWorld)

The Church-Turing thesis is really the Church-Turing *hypo*thesis, because it hasn't ever been "proven". It is less of a mathematical therom and more of a statement of faith.

- Are you a believer? If so, devise a cult to convert a billion people into the religion.

4 The Halting Problem

> Mathematics, rightly viewed, possesses not only truth, but supreme beauty – a beauty cold and austere ... without appeal to ... our weaker nature ... sublimely pure ... capable of a stern perfection... Real life is ... a long second-best, a perpetual compromise between the ideal and the possible; but the world of pure reason knows no compromise, no practical limitations, no barrier to the creative activity ... [it is] where ... our nobler impulses can escape from the dreary exile of the actual world.
>
> (Bertrand Russell)

The neurotic Russell wanted a perfect refuge, a mathematics free of contradiction, and spent years writing *Principia Mathematica* to create this fortress. His dream was shattered forever with Gödel's Incompleteness Theorem. And the Halting Problem relies on exactly the same idea.

The theorem has spawned a host of different interpretations. The philosopher J. R. Lucas said, "Gödel's theorem seems to me to prove that Mechanism is false, that is, that minds cannot be explained as machines."[1] This philosophical position is called Mechanism. Hofstadter mentions that this argument was a major motivation for him to write *Gödel, Escher, Bach*, though he disagrees with it.[2] Physicist and philosopher Roger Penrose, another Mechanist, says because Gödel showed the mind is not a machine, there must be something kind of mystical and immaterial in the brain that causes consciousness, and for some reason points to quantum microtubules[3]. Quantum computer scientist Scott Aaronson gets a lot of mileage out of making fun of Penrose.[4] The developmental psychologist and philosopher Jean Piaget had to change his entire theory of child development to accommodate for Gödel's discovery.[5]

But for me, the flavor of the problem is best captured in the following thought experiment.

> [Newcomb's Paradox.] Suppose that a super-intelligent Predictor shows you two boxes: the first box has $1,000, while the second box has either $1,000,000 or nothing. You don't know which is the case, but the Predictor has already made the choice and either put the money in or left the second box empty. You, the Chooser, have two choices: you can either take the second box only, or both boxes. Your goal, of course, is money and not understanding the universe.
>
> Here's the thing: the Predictor made a prediction about your choice before the game started. If the Predictor predicted you'll take only the second box, then he put $1,000,000 in it. If he predicted you'll take both boxes, then he left the second box empty. The Predictor has played this game thousands of times before, with thousands of people, and has never once been wrong. Every single time someone picked the second box, they found a million dollars in it. Every single time someone took both boxes, the found that the second box was empty.
>
> First question: Why is it obvious that you should take both boxes? Right: because whatever's in the second box, you'll get $1,000 more by taking both boxes. The decision of what to put in the second box has already been made; your taking both boxes can't possibly affect it.
>
> Second question: Why is it obvious that you should take only the second box? Right: because the Predictor's never been wrong! Again and again you've seen one-boxers walk away with $1,000,000, and two-boxers walk away with only $1,000. Why should this time be any different?
>
> Q: How good is the Predictor's computer?
>
> Scott: Well, clearly it's pretty good, given that he's never been wrong. We're going to get to that later.
>
> This paradox was popularized by a philosopher named Robert Nozick in 1969. There's a famous line from his paper about it: "To almost everyone, it is perfectly clear and obvious what should be done. The difficulty is that these people seem to divide almost evenly on the problem, with large numbers thinking that the opposing half is just being silly."
>
> There's actually a third position—a boring "Wittgenstein" position—which says that the problem is simply incoherent, like asking about the unstoppable force that hits the immovable object. If the Predictor actually existed, then you wouldn't have the freedom to make a choice in the first place; in other words, the very fact that you're debating which choice to make implies that the Predictor can't exist.
>
> (Scott Aaronson, *Quantum Computing Since Democritus)*

- Which box do you take and why?

[1] *Minds, Machines, and Gödel*, 1959
[2] p466, *GEB*
[3] *Shadows of the Mind*, 1994
[4] https://www.scottaaronson.com/democritus/lec10.5.html; https://www.scottaaronson.com/writings/captcha.html
[5] Piaget's Neo-Gödelian Turn, http://journals.sagepub.com/doi/abs/10.1177/0959354316672595?journalCode=tapa, 2016