Simulating consciousness and the quantum brain

Ieva Cepaite investigates the fundamental nature of consciousness

Every once in a while I find myself slumped against a car door, cheek pressed to the cool glass and staring out at the grey of the moving landscape - I mean, it's Scotland, it's usually grey - as my brain inevitably spirals down a philosophical wormhole: are we living in a simulation? Is anything real? Was that a cow or just a really impressive stack of hay?

It's nice to sometimes end the whole thing with a familiar and comforting argument - *Cogito ergo sum* - "I think therefore I am." I mean, who am I to question a guy like Descartes, who has obviously done most of the hard work for me, right? It's the kind of statement that retains just enough mystery but gives the whole endless spiral of doubt a bit of closure. It's sensible. But here's the thing: what if *I'm* a computer simulation too? Am I really thinking, then, or is it just a series of commands issued by an overarching and extremely complex bit of circuitry? Moreover, is it possible *for us* to simulate humans?

Alan Turing, for example, wrote entire papers concerning this topic. It's where the eponymous Turing Test initially came into being - if a computer interacts with us in a way that is indistinguishable from a human, then of course we could say the computer isn't 'really' thinking, that it's just a simulation. But on the same grounds, we could also say that other people aren't really thinking, that they merely act as if they're thinking.

Of course, there are currently plenty of tasks that a human can do but a computer struggles with. This is the idea behind things like CAPTCHA, the tests used to distinguish real users from spambots online. The key property of these tests is that a computer should be able to generate and grade them, but not pass them. Only humans should be able to pass the tests. So basically, they capitalise on the *failures* of artifical intelligence (AI).

In trying to write programs to simulate human intelligence, we're competing against *a billion years of evolution.* One counterintuitive consequence is that it's much easier to program a computer to beat Garry Kasparov at chess than to program a computer to recognize faces under varied lighting conditions. Often the hardest tasks for AI are the ones that are trivial for a five-year-old since those are the ones that are so hardwired by evolution that we don't even think about them.

Illustration by Prerna Vohra

Often the hardest tasks for AI are the ones that are trivial for a five-year-old since those are the ones that are so hardwired by evolution that we don't even think about them

Then again, the above is not an argument against the possibility of writing a program that's complex enough to capture the nuance of image recognition or other tasks. It may currently be unfeasible, but nothing tells us it's *impossible.* There is, however, a slightly different argument against the equivalence of machines and humans - *reproducing* human behaviour is not the same as *understanding* something you do.

This is illustrated by a thought experiment, first proposed around 1980 called *Searle's Chinese Room.* Let's say you don't speak Chinese. You sit in a room, someone passes you paper slips through a hole in the wall with questions written in Chinese, and you're able to answer the questions (again in Chinese) just by consulting a rule book. In this case, you might be carrying out an intelligent Chinese conversation; yet by assumption, you don't understand a word of Chinese. Therefore, symbol-manipulation can't produce understanding.

Over the years there have been a lot of holes poked in this argument. For example, you might not understand Chinese, but the rule book does! Or if you like, understanding Chinese is an emergent property of the system consisting of you and the rule book, in the same sense that understanding English is an emergent property of the neurons in your brain. We are also not considering the complexity of the system itself - how large is the rule book and what's the method to look up the required parts of it to carry out an intelligible conversation? These are likely so outlandish and complex that some form of insight or understanding may be a requirement for it to actually function.

Everyone who talks about this stuff is really tiptoeing around the question of consciousness. See, consciousness has these weird twin properties. On the one hand, it's arguably the most mysterious thing we know about, and on the other, not only are we directly aware of it, but in some sense it's the *only* thing we're directly aware of. And granting consciousness to a robot seems strangely equivalent to *denying that one is conscious oneself.*

If a computer interacts with us in a way that is indistinguishable from a human, then of course we could say the computer isn't "really" thinking... But we could also say that other people aren't really thinking, that they merely act as if they're thinking

One of my favourite ways around this is one proposed by the philosopher David Chalmers: if computers someday *emulate* humans in every observable respect, then we'll be compelled to regard them as conscious, for exactly the same reasons we regard other *people* as conscious. And as for *how* they could be conscious - well, we'll understand that just as well or as poorly as we understand how a bundle of neurons could be conscious. Yes, it's mysterious, but the one mystery doesn't seem so different from the other.

This leads us to the final chapter of this convoluted journey and to the musings of a man named Roger Penrose, arguably one of the most famous living mathematical physicists in the world. He doesn't believe in the possibility of AI in the sense that we've discussed thus far and while his arguments for this are fairly complex, they boil down to the idea that there is something inherently quantum mechanical about our consciousness which a computer simply cannot imitate.

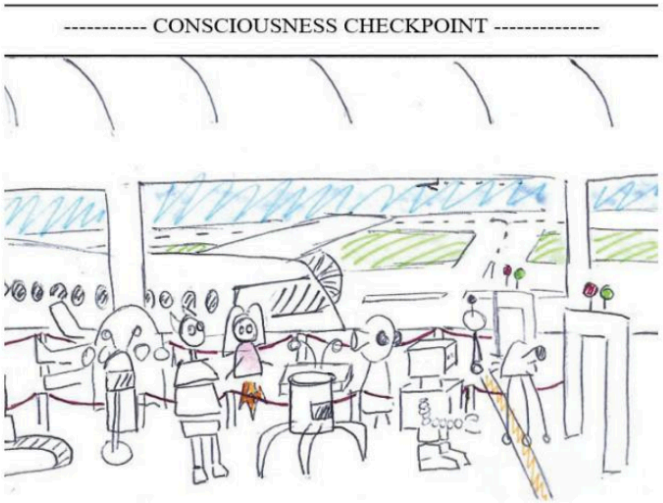

Illustration by Prerna Vohra

In order to illustrate why he believes this, Penrose presents a brilliant thought experiment - the teleportation machine. This is a machine that whisks you around the galaxy at the speed of light, by simply scanning your whole body, encoding all the cellular structures as pure information, and then transmitting the information as radio waves. When the information arrives at its destination, nanobots use the information to reconstruct your physical body down to the smallest detail. Oh, and since it's weird to have two of you running around, the first copy of you is killed. Sounds appealing? Are you going to tell me you're somehow attached to the *particular atoms* that currently reside in your brain? As long as the information is safely on its way to Mars, who cares about the original meat hard drive? Those atoms are replaced every few weeks anyway, so it *can't* be the atoms themselves that make you you; it has to be the patterns of information they encode, patterns that aren't obvious to decode. Penrose suggests that they are quantum in nature.

There is a way out of the above problem though, one which is loosely based around the field of quantum computing -

Suppose some of the information that made you was actually *quantum* information, which exhibits strange things like entanglement and superposition that make no physical sense on a larger scale. As a consequence of something called the No-Cloning Theorem (a name I feel is fairly self-explanatory) which is specific to quantum mechanics, no such machine could possibly work as claimed.

This is not to say that you couldn't be teleported around at the speed of light. But the teleportation process would have to be very different from the one above - it could *not* involve copying you and then killing the original copy. Either you could be sent as quantum information, or else - if that wasn't practical - you could use the famous quantum teleportation protocol, which sends only classical information, but also requires prior quantum entanglement between the sender and the receiver (a tough thing to do with an entire human being's worth of information). In either case, the original copy of you would disappear unavoidably, as part of the teleportation process itself.

So we're left with the same mystery we began with, if maybe with a little more food for thought. Perhaps someday we'll be able to discover and explain the quantum mechanics of the brain, but until then all we can do is train a computer to play chess or spam twitter.

Ieva Cepaite is a 5th Year Computational Physics student with a keen interest in Quantum Computing