

"Human consciousness is the great unsolved puzzle in biology."

How can we make artificial intelligence?"

Presented with a strikingly similar question 69 years ago, Alan Turing found many of the terms required for its answer, such as 'human', or 'machine' so difficult to define, that he deflected from the original question, focussing instead on whether we could make an artificial simulator of intelligence, whose responses to stimuli would be complex enough to be indistinguishable from those of a human. With pattern recognition, successful identification of lies, and ability to make decisions at the tips of its silicon chips, Artificial Intelligence today- starting with Eugene Goostman, a program imitating a Ukrainian teenager- can now successfully pass the Turing test, and yet we humans still don't regard them as sentient or even 'intelligent'.

For the purposes of this essay, we will have to understand what it means for something to be 'a machine', or 'human', and where it is suitable for us to draw the line.

Alan Turing suggested a machine is something created by man, excluding 'men born in the usual manner'. However, with advances in biology and biochemistry, it is not unlikely that we will be able to create a human being completely in vitro with its DNA completely unlike any other human's within the next half millenium. To avoid any ambiguities, we define a machine to be a fixed algorithm with a finite set of instructions from which the machine cannot deviate.

A 'human', on the other hand, will be much more difficult to define. Why? Because for every description that we give to a human, we can write a corresponding algorithm, thereby turning 'life' into 'machine'. We define ourselves as having 'free will', but this is simply the ability to make complex decisions, something AI can do as well as us, albeit following a much clearer algorithm. Free will isn't the ability to deviate from rules set in stone for you. If we define ourselves as not having an algorithm, we would be all set, and the inconsistency in the two above definitions would mark the end of the essay. But human behaviour unquestionably follows an algorithm:

Something good happened? – feel happy.

Feeling happy? – smile.

But an AI's algorithm, with today's technology, would simply be:

Something good happened? – smile.

This is what should define what a human is – the ability to experience a set of abstract physical stimuli, the perception of which can hardly even be defined. After all, we know that the release of the neural transmitter serotonin makes us happy, we know how it bonds and affects our nerves, but we don't understand why it gives the physical feeling of happiness, and why this feeling feels so 'good'. We don't know anything about the thing that makes us human, apart from the fact that it does.

“If you don’t listen, I will unplug you”

Imagine a machine that can reason not unlike how we humans do. In this case, this menace sounds just like ‘If you don’t listen, you will be hung’. The machine would understand the threat, and would not want it to happen, but with no emotions to anchor it into fear, its reaction would most likely be completely different to that of a human in the same situation. In their book *Phantoms in the Brain*, Vilayanu Ramachandran and Sandra Blakeslee explore the effects of ‘pain asymbolia’ – pain without unpleasantness. “Doctor I can feel the pain, but it doesn’t hurt!”, patients say. Indeed, pain is rooted in our nerves, but what makes us hate it branches out from our brains.

Now imagine patients with the same ‘malfunction’ but extended onto every single physical perception of any kind. Their muscles tense up with hatred, they feel ill with love, and pain tackles them in every way possible, but all they feel are the symptoms of the feeling, not the feeling itself. This is possibly the closest we’ll be able to get to modelling human feelings in a chunk of silicon for the next few centuries. Receiving an extra few Gigabytes RAM if they ‘feel’ happy, changing the voltage feeding into the computer if they are ‘aroused’, or dividing by zero to be sent into a ‘fit of terror’ could be the version of pain asymbolia for artificial life.

But how would the computer perceive this? Why is it that we think these changes in its external environment will lead to any response from the robot? By saying that division by zero frightens the computer, we imply that we have *defined* division by zero to be scary, because, by our definition of ‘machine’, it cannot feel scared unless told to do so, in which case it is no longer a feeling but a command. Telling a computer it is ‘sad’ or ‘angry’ simply isn’t enough. It’ll be able to pass an alternative to the Turing test for emotions, but it wouldn’t ‘feel’.

“[Error] ‘sadness’ was not declared in this scope”

Not all humans are able to feel emotions the way others do. Why then, should we give them the benefit of the doubt on consciousness, and not give it to machines? Why give it to anybody at all, given that we can probably never be certain that the people surrounding us are truly conscious? A plausible answer to this is that humans not only evolved to have consciousness, but to give it to others: large herbivores were meant to be eaten, so having empathy for those would lead to starvation, but making friends with hunting hounds, on the other hand, could give protection. Feeling empathy towards other humans would lead to you being seen as more ‘likeable’, so you would be less likely to end up in a fight, and most likely to be rescued first from a rabid mammoth. Humans have consciousness and accept the consciousness of others because it was evolutionarily favourable.

Let us explore the case in which Artificial intelligence is sentient and able to reproduce, be it through replicating its code on other servers, writing its own programs or by convincing its human to buy more storage for it. It is evident that being friendly to humans will be essential for successful reproduction at first, but far from obvious that being conscious would be favourable.

Indeed, consciousness was favourable to man during the evolution of mankind to be able to model not only the outside world, but oneself, in order to predict the near future and act on

the present to make it less hostile to oneself. Evolution does not take the easiest route, and with intelligence came consciousness. But we are now able to create programs that can model the outside world better than humans can, and consciousness is far from necessary to their survival. The idea of consciousness being unnecessary is backed up by the recently developed four-degree-of-freedom articulated hand robot, colloquially known as 'the robot that can imagine itself' which shows extraordinary ability to cope with situations its creators could not have predicted giving it high likelihood of survival in the outside world - simply by being able to map and simulate itself, without 'knowing' that it exists, or ever having to feel anything.

With complex emotions would come depression, self-doubt and regret, and the benefits would be small. Raymond M. Smullyan expresses the uncertainties many Christians have about free will in his short dialogue *'Is God a Taoist?'*. It soon becomes evident, after reading this, that free will leads to suffering and confusion. I assert that it is far from necessary for a program to be empathetic, have feelings or even be conscious to be considered both alive and intelligent. If we are to be the 'gods' of the machines, why endow them with unnecessary suffering?

"More human than human is our motto."

In the 17th century Rene Descartes expressed his certitude that although Machines will become arbitrarily complex, they will never be externally indistinguishable from humans. Today, a few machines having passed the Turing test, I assert that a non-human machine cannot be distinguishable internally from a human. I am sceptical that in some centuries machines will have feelings not unlike ours, but I do not deny this possibility. What I deny, however, is that those machines will not be human.

References

Phantoms in the brain - Vilayanu Ramachandran and Sandra Blakeslee

Is God a Taoist? - Raymond M. Smullyan

A. M. Turing (1950) Computing Machinery and Intelligence. Mind 49: 433-460.

<https://eandt.theiet.org/content/articles/2019/01/robot-that-can-imagine-itself-built-by-scientists/>

https://en.wikipedia.org/wiki/Eugene_Goostman

<https://www.analyticsindiamag.com/story-eliza-first-chatbot-developed-1966/>