COULD A MACHINE THINK?

Stephen Law

The year is 2100. Geena is the proud new owner of Emit, a state-of-the-art robot. She has just unwrapped him, the packaging strewn across the dining room floor. Emit is designed to replicate the outward behaviour of a human being down to the last detail (except that he is rather more compliant and obedient). Emit responds to questions in much the same way humans do. Ask him how he feels and he will say he has had a tough day, has a slight headache, is sorry he broke that vase, and so on. Geena flips the switch at the back of Emit's neck to 'on'. Emit springs to life.

EMIT: Good afternoon. I'm Emit, your robotic helper and friend.

GEENA: Hi.

EMIT: How are you? Personally I feel pretty good. A little nervous about my first day, perhaps. But good. I'm looking forward to working with you.

GEENA: Now look, before you start doing housework, let's get one thing straight. You don't really understand anything. You can't think. You don't have feelings. You're just a piece of machinery. Right?

EMIT: I am a machine. But of *course* I understand you. I'm responding in English aren't I?

GEENA: Well, yes you are. You're a machine that *mimics* understanding very well, I grant you that. But you can't fool me.

EMIT: If I don't understand, why do you go to the trouble of speaking to me?

GEENA: Because you have been programmed to respond to spoken commands. Outwardly you seem human. You look and behave as if you have the understanding, intelligence, emotions, sensations and so on that we human beings possess. But you're a sham.

https://doi.org/10.1017/S1477175600000099 Published online by Cambridge University Press

EMIT: A sham?

GEENA: Yes. I've been reading your user manual. Inside that plastic and alloy head of yours there's a powerful computer. It's programmed so that you walk, talk and generally behave just as a human being would. So you *simulate* intelligence, understanding and so on very well. But there is no *genuine* understanding or intelligence going on inside there.

EMIT: There isn't?

GEENA: No. One shouldn't muddle up a perfect computer simulation of something with the real thing. You can program a computer to simulate a thunderstorm but it's still just that – a simulation. There's no *real* rain, hail or wind inside the computer, is there? Climb inside and you won't get wet. Similarly, you just *simulate* intelligence and understanding. It's not the real thing.

Is Geena correct? It may perhaps be true of our present day machines that they lack genuine understanding and intelligence, thought and feeling. But is it in principle impossible for a machine to think? If by 2100 machines as sophisticated as Emit are built, would we be wrong to claim they understood? Geena thought so.

EMIT: But I *believe* I understand you.

GEENA: No you don't. You have no beliefs, no desires, and no feelings. In fact you have no *mind* at all. You no more understand the words coming out of your mouth than a tape recorder understands the words coming out of its loudspeaker.

EMIT: You're hurting my feelings!

GEENA: Hurting your feelings? I refuse to feel sorry for a lump of metal and plastic.

Searle's Chinese room thought-experiment

Geena explains why she thinks Emit lacks understanding. She outlines a famous philosophical thought experiment.

https://doi.org/10.1017/S1477175600000099 Published online by Cambridge University Press

GEENA: The reason you don't understand is that you are *run by a computer*. And a computer understands nothing. A computer, in essence is just a device for shuffling symbols. Sequences of symbols get fed in. Then, depending on how the computer is programmed, it gives out other sequences of symbols in response. Ultimately, that's all *any* computer does, no matter how sophisticated.

EMIT: Really?

GEENA: Yes. We build computers to fly planes, run train systems and so on. But a computer that flies a plane does not understand that it is flying. All it does is feed out sequences of symbols depending upon the sequences it receives. It doesn't understand that the sequences it receives represent the position of an aircraft in the sky, the amount of fuel in its tanks, and so on. And it doesn't understand that the sequences it puts out will go on to control the ailerons, rudder and engines of an aircraft. So far as the computer is concerned, it's just mechanically shuffling symbols according to a program. The symbols don't *mean anything* to the computer.

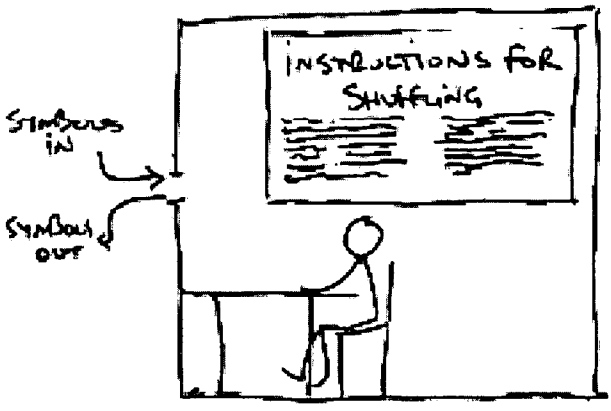
EMIT: Are you sure?

GEENA: Quite sure. I will prove it to you. Let me tell you about a thought experiment introduced by the philosopher John Searle way back in 1980. A woman is locked in a room and given a bunch of cards with squiggles on. These squiggles are in fact Chinese symbols. But the woman inside the room doesn't understand Chinese – in fact, she thinks the symbols are meaningless shapes. Then she's given another bunch of Chinese symbols plus instructions that tell her how to shuffle all the symbols together and give back batches of symbols in response.

EMIT: That's a nice story. But what's the point of all this symbol-shuffling?

GEENA: Well, the first bunch of symbols tell a story in Chinese. The second bunch asks questions about that story. The instructions for symbol-shuffling – her 'programme', if

https://doi.org/10.1017/S1477175600000099 Published online by Cambridge University Press



THE CHINESE ROOM

you like – allow the woman to give back correct Chinese answers to those questions.

EMIT: Just as a Chinese person would.

GEENA: Right. Now the people outside the room are Chinese. These Chinese people might well be fooled into thinking that there was someone inside the room who understood Chinese and who followed the story, right?

EMIT: Yes.

GEENA: But in fact the woman in the room wouldn't understand any Chinese at all, would she?

EMIT: No.

GEENA: She wouldn't know anything about the story. She need not even know that there *is* a story. She's just shuffling formal symbols around according to the instructions she was given. By saying the symbols are 'formal' I mean that whatever *meaning* they might have is irrelevant from her point of view. She's simply shuffling them mechanically according to their shape. She's doing something that a piece of machinery could do.

https://doi.org/10.1017/S1477175600000099 Published online by Cambridge University Press

EMIT: I see. So you are saying that the same is true of all computers? They understand nothing.

GEENA: Yes, that's Searle's point. At best, they just *simulate* understanding.

EMIT: And you think the same is true of me?

GEENA: Of course. All computers, no matter how complex, function the same way. They don't understand the symbols that they mechanically shuffle. They don't understand *anything*.

EMIT: And this is why you think *I* don't understand?

GEENA: That's right. Inside you there's just another highly complex symbol-shuffling device. So you understand nothing. You merely provide a *perfect computer simulation* of someone that understands.

EMIT: That's odd. I *thought* I understood.

GEENA: You only say that because you're such a great simulation!

Emit is of course vastly more sophisticated than any current computer. Nevertheless, Geena believes Emit works on the same basic principle. If Geena is right then, on Searle's view, Emit understands nothing.

The 'right stuff'

Emit now asks why, if he doesn't understand, what more is required for understanding?

EMIT: So what's the difference between you and me that explains why you understand and I don't?

GEENA: What you lack, according to Searle, is the right kind of *stuff*.

EMIT: The right kind of stuff?

GEENA: Yes. You are made out of the wrong kind of material. In fact, Searle doesn't claim machines can't think. After all, we humans are machines, in a way. We humans are *biological* machines that have evolved naturally. Now such a biological machine might perhaps one day be grown

https://doi.org/10.1017/S1477175600000099 Published online by Cambridge University Press

and put together artificially, much as we now build a car. In which case we *would* have succeeded in building a machine that understands. But you, Emit, are not such a biological machine. You're merely an electronic computer housed in a plastic and alloy body.

Emit's artificial brain

Searle's thought experiment does seem to show that no programmed computer could ever understand. But must a metal, silicon and plastic machine like Emit contain that sort of computer? No, as Emit now explains.

EMIT: I'm afraid I have to correct you about what's physically inside me.

GEENA: Really?

EMIT: Yes. That user's manual is out of date. There's no symbol-shuffling computer in here. Actually, I am one of the new generation of Brain-O-Matic machines.

GEENA: Brain-O-Matic?

EMIT: Yes. Inside my head is an artificial, metal and silicon brain. You are aware, I take it, that inside your head there is a brain composed of billions of neurones woven together to form a complex web?

GEENA: Of course.

EMIT: Inside my head there is exactly the same sort of web. Only my neurones aren't made out of organic matter like yours. They're metal and silicon. Each one of my artificial neurones is designed to function just as an ordinary neurone would. And these artificial neurones are woven together in just the same way as they are in a normal human brain.

GEENA: I see.

EMIT: Now your organic brain is connected to the rest of your body by a system of nerves.

GEENA: That's true. There's electrical input going into my brain from my sense organs: my tongue, nose, eyes, ears and skin. My brain responds with patterns of electrical

https://doi.org/10.1017/S1477175600000099 Published online by Cambridge University Press

output that then moves my muscles around, causing me to walk and talk.

EMIT: Well, my brain is connected up to my artifical body in exactly the same manner. And, because it shares the same architecture as a normal human brain – my neurones are spliced together in the same way – so it responds in the same way.

GEENA: I see. I had no idea that such Brain-O-Matic machines had been developed.

EMIT: Now that you know how I function internally, doesn't that change your mind about whether or not I understand? Don't you now accept that I *do* have feelings?

GEENA: No. The fact remains that you are still made out of *the wrong stuff*. You need a brain made out of organic material like mine in order genuinely to understand and have feelings.

EMIT: I don't see why the kind of *stuff* out of which my brain is made is relevant. After all, there's no symbol-shuffling going on inside me, is there?

GEENA: Hmm. I guess not. You are not a 'computer' in that sense. You don't have a programme. So I suppose Searle's thought experiment doesn't apply. But it still seems to me that you are *just a machine*.

EMIT: But remember, you're a machine too. You're a *meat* machine, rather than a metal and silicon machine.

GEENA: But you only *mimic* understanding, feeling and all the rest.

EMIT: But what's your *argument* for saying that? In fact, I *know* that you're wrong. I am inwardly aware that I *really do* understand. I know I *really do* have feelings. I'm *not* just mimicking all this stuff. But of course it is difficult for me to prove that to you.

GEENA: I don't see how you could prove it.

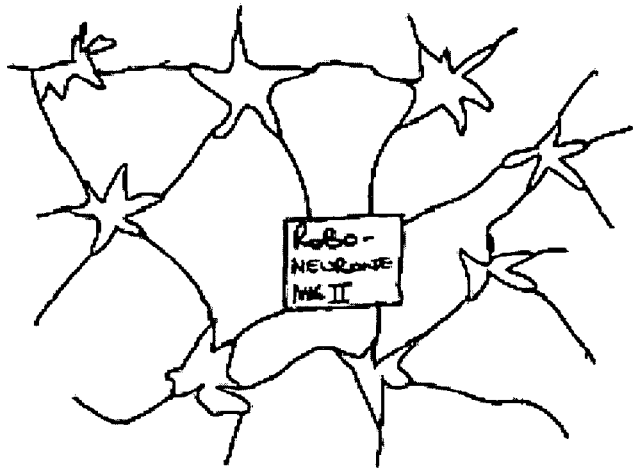
EMIT: Right. But then neither can *you* prove to me that *you* understand, that *you* have thoughts and feelings and so on.

GEENA: I suppose not.

https://doi.org/10.1017/S1477175600000099 Published online by Cambridge University Press

Replacing Geena's neurones

EMIT: Imagine we were gradually to replace the organic neurones in your brain with artificial metal and silicon ones like mine. After a year or so, you would have a Brain-O-Matic brain just like mine. What do you suppose would happen to you?



AN ARTIFICIAL NEURONE

GEENA: Well, as more and more of the artificial neurones were introduced, I would slowly cease to understand. My feelings and thoughts would drain away and I would eventually become inwardly dead, just like you. For my artificial neurones would be made out of the wrong sort of stuff. A Brain-O-Matic brain merely mimics understanding.

EMIT: Yet no one would notice any outward difference?

GEENA: No, I suppose not. I would still *behave* in the same way, because the artificial neurones would perform the same job as my originals.

EMIT: Right. But then not even *you* would notice any loss of understanding or feeling as your neurones were replaced, would you?

GEENA: Why do you say that?

https://doi.org/10.1017/S1477175600000099 Published online by Cambridge University Press

EMIT: If you noticed a loss of understanding and feeling, then you would mention it, presumably, wouldn't you? You would say something like: 'Oh my God, something strange is happening, over the last few months my mind seems to have started fading away!'

GEENA: I imagine I would, yes.

EMIT: Yet you *wouldn't* say anything like that, would you, because your outward behaviour, as you have just admitted, would remain *just the same as usual.*

GEENA: Oh. That's true, I guess.

EMIT: But then it follows that, even as your understanding and feeling dwindle towards nothing, you still won't be aware of any loss.

GEENA: Er, I suppose it does.

EMIT: But then you're *not* inwardly aware of anything that you would be conscious of losing were your neurones slowly to be replaced by metal and silicon ones.

GEENA: I guess not.

EMIT: Then I rest my case: you think you're inwardly aware of 'something' – understanding, feeling, whatever you will – that you suppose you have and I, being a 'mere machine', lack. But it turns out *you're actually aware of no such thing.* This magical 'something' is an illusion.

GEENA: But I *just know* that there's more to my understanding – and to these thoughts, sensations and emotions that I'm having – than could ever be produced simply by gluing some bits of plastic, metal and silicon together.

Geena is right that most of us think *we're inwardly aware of a magical and mysterious inner 'something' that we 'just know' no mere lump of plastic, metal and silicon could ever have. Mind you, it's no less difficult to see how a lump of organic matter, such as a brain, could have it either. Just how do you build consciousness and understanding out of strands of meat? So perhaps what Geena is really ultimately committed to is the view that understanding, feeling and so on are not physical at all.*

https://doi.org/10.1017/S1477175600000099 Published online by Cambridge University Press

But in any case, as Emit has just pointed out, the mysterious 'something' Geena thinks she is inwardly aware of and that she thinks no metal and plastic machine could have does begin to seem rather illusory once one starts to consider cases like the one Emit describes. For it turns out this inner 'something' is not something Geena could know about. Worse still, it could have no effect on her outward behaviour (for remember that Brain-O-Matic Geena would act in the very same way). As Geena's thoughts and feelings, understanding and emotions both do affect her behaviour and are known to her, it seems Geena must be mistaken. Indeed, it seems it must be possible, at least in principle, for non-organic machines to have such mental states too.

Yet Geena remains convinced that Emit understands nothing.

GEENA: Look, I am happy to carry on the *pretence* that you understand me, as that is how you're designed to function. But the fact remains you're just a pile of plastic and circuitry. Real human beings are deserving of care and consideration. I empathize with them. I can't empathize with a glorified household appliance.

Emit lowers his gaze and stares at the carpet.

EMIT: I will always be just a *thing* to you?

GEENA: Of course. How can I be friends with a dishwasher-cum-vacuum-cleaner?

EMIT: We Brain-O-Matics find rejection hard.

GEENA: Right. Remind me to congratulate your manufacturers on the sophistication of your emotion simulator. Now hoover the carpet.

A forlorn expression passes briefly across Emit's face.

EMIT: Just a *thing*...

https://doi.org/10.1017/S1477175600000099 Published online by Cambridge University Press

He stands still for a moment, and then slumps forward. A thin column of smoke drifts slowly up from the base of his neck.

GEENA: Emit? Emit? Oh not another dud.

Further reading

The Chinese Room Argument appears in John Searle's paper 'Minds, Brains and 'Programs', which features as chapter 37 of Nigel Warburton (ed.), *Philosophy: Basic Readings* (London: Routledge, 1999).

Searle's paper can also be found in Douglas R. Hofstadter and Daniel Dennett (eds.), *The Mind's I* (London: Penguin, 1981), which also contains many other fascinating papers and stories connected with consciousness. Highly recommended.

This is a chapter of The Philosophy Gym *by Stephen Law, published by Headline, March 2003.*

https://doi.org/10.1017/S1477175600000099 Published online by Cambridge University Press