I think, therefore…

Published on 2 February 2009

By Christine Evans-Pughe

E&T reports on the latest developments in building a conscious robot – one of the biggest modern challenges of science and technology.

For the last few years, a group of British scientists, led by Professor Owen Holland of the University of Essex, has been engaged in an ambitious project to build a conscious robot. They officially finished in 2007 and yet, disappointingly, no conspicuously conscious form of machine life has emerged from their labs.

However, Prof Holland has now been awarded a €3m EU grant to develop the concept further in a three-year project called 'ECCEROBOT' (embodied cognition in a compliantly engineered ROBOT). But how far did they get with the first project, and what can we expect from the next one?

With their initial £500k grant from the Engineering and Physical Sciences Research Council, Prof Holland, and Professor Tom Troscianko and Dr Iain Gilchrist of the University of Bristol had set out to explore the ideas put forward by Thomas Metzinger, Antonio Damasio and others that, to be conscious, an intelligent agent needs an internal model of itself that can interact with an internal model of the world.

It was vital, they felt, to have a robot with an internal model similar to our own, and this meant building a robot with a body like our own. So they built a human-sized and human-like robot (called CRONOS) made of artificial bones, tendons and muscles, and linked it to a detailed computer simulation of itself called SIMNOS.

By driving CRONOS and SIMNOS with a visual system modelled on the human brain, Prof Holland's team hoped to study what happened when SIMNOS started to 'imagine' actions autonomously and carry them out using the real robot body.

As a way of assessing progress, Prof Holland's former PhD student Dr David Gamez developed a spiking neural network simulator (one that mimics the short and sudden increases in voltage – spikes – that biological neurons use to send information) for controlling SIMNOS' eye muscle movements. The idea was to examine the simulator for signs of consciousness and use the neural network to control CRONOS.

Dr Gamez has managed to show that different parts of the neural network were predicted to be conscious according to Tononi's, Aleksander's and Metzinger's theories, but it was not possible to predict the absolute amount of consciousness because the measures had not been calibrated on normal waking human subjects.

"This work is at an extremely early stage and a great deal of research is needed to improve the accuracy of our predictions," writes Dr Gamez. "It is hoped that this will eventually lead to a more systematic science of consciousness that includes both natural and artificial systems within a single conceptual and experimental framework."

As Prof Holland explains, the project proved every bit as difficult as they thought. "Most of the time was taken up building the infrastructure. We had to create the robot, the internal model, and a way of getting the robot to look at the world, put things into its internal model and be able manipulate the model and make decisions.

"When we got to that stage, we hoped we would – in a sense – be able to turn it loose and see what it did and what it preferred doing. In fact, we have only just reached that stage a year after the official end of the project."

Philosophical vagueness

It's not surprising that Prof Holland and his colleagues didn't meet their goal in one short project.

Philosophers have been arguing over tricky conceptual questions to do with consciousness – such as what is access consciousness versus phenomenal consciousness, and are there zombies – for centuries, without coming up with any decisive answers.

Neuroscientists who are tooled-up with the brain scanners that show changes in the brain's blood flow or neurons' electrical or

magnetic activity are shining more light into this philosophical vagueness, but there is no agreement about what consciousness is, what is caused by consciousness or what causes it, according to Professor Geraint Rees, one of the leading scientists working on the neural basis of consciousness.

Prof Rees suggests that a simple working definition of consciousness might be: level of consciousness (being awake rather than being asleep), contents of consciousness (tasting a glass of wine versus smelling a rose), and self-consciousness (I think therefore I am). His research group at the Institute of Cognitive Neuroscience and Wellcome Trust Centre for Neuroimaging at University College London looks at sensory perception and vision and uses brain imaging to understand what patterns of activity are associated with the contents of consciousness as distinct from the unconscious.

He explains the difference: "If one is at a wine-tasting, for instance, there must be some basic underlying computing going on – your tongue has receptors, they are activated in a particular way by the fine Bordeaux versus the cheap Chardonnay, and somehow there is some processing in your brain giving rise to the 'Ooh that's nice!' response.

"But, however hard you think, you can't get down to that basic underlying computation. That's the distinction between the conscious contents – 'your thoughts' – and the unconscious sea on which they are floating and on which they rely."

One of the broad findings that Prof Rees' group and others (such as Wolf Singer's at the Max Planck Institute in Germany, and Christof Koch's at the California Institute of Technology in America) have made is that unconscious information permeates most of the brain. "The classic model is that as we start processing the information coming in from the outside, most of that processing at early stages is unconscious, but then by the time it gets to high-level information, such as recognising a friend, the information is all conscious," explains Prof Rees.

"But our findings suggest that isn't the case. We can identify neural activity associated with conscious representations already at very early levels of processing; and activity associated with unconscious representations at high levels. So something else must distinguish conscious and unconscious processing, not location within the brain."

The second set of findings is that other areas of the brain as well as the visual cortex are involved when we become consciously aware of things. "So it isn't only the neurons that respond to the visual world but also neurons in the parietal and prefrontal cortex of the brain that have to be activated," says Prof Rees.

Prof Rees and his team have also been making progress on correlating brain activity with specific thoughts. "Using a process called binocular rivalry in which perception alternates between different images presented to each eye, we look at the pattern of activity in the brain and predict whether the subject is perceiving one image or the other.

"In such a situation, we can now very accurately predict what a subject is consciously perceiving, so we're trying to generalise that to other situations," he says. "It's a building block to build a more general mind-reading device. It's a good few decades before we get anything more sophisticated but it's another way of looking at the problem from a practical point of view."

These developments are improving our understanding of consciousness in humans but don't get us a great deal further towards detecting it in a machine. As Prof Rees says: "It's deeply difficult to establish whether any robotic system has anything corresponding to what we regard as consciousness."

We discover that people are conscious by asking them to report their non-observable internal states, although a human can lie, so their internal state can be quite different to what they report. A car has non-observable internal states – for example its electronic ignition system has a crude internal representation of the world and it communicates to me through a blinking light panel – but do we really want to say the car is conscious? And how about a computer simulation of consciousness?

"There are supercomputers that undertake very detailed simulation of the weather but no-one actually thinks the weather is IN the computer! Why would we expect, if we ran a model of consciousness on a computer, that the computer would be conscious?" asks Prof Rees. "I'm not saying you can't and it isn't but it's a layered, rich and difficult question."

A society of brains

Another way of looking at consciousness is to consider brains not as isolated organs but members of a society of brains that are firmly embedded in a social culture and environment.

Professor Wolf Singer, director of the Max Planck Institute for Brain Research in Frankfurt, believes that we will only get close to bridging the gap between the subjective experience and neuronal processes if we start to consider the brain as a social organ.

"What we know about ourselves is dependent on early imprinting and development in a social and cultural network. The human brain develops over at least 20 years and there is a lot of pruning of connections – and through this epigenetic shaping process, a lot of instructions enter the brain that go well beyond the genetically determined architecture," he says.

Singer's hunch is that a conscious machine might only result if you raised it in a social cultural environment like ours. But would it be able to feel pain and pleasure? "You can build machines that follow non-linear dynamics and endow them with evaluating systems and value-assigning systems that would give a red light when they 'feel' there is a problem and a green light when they have a solution, whether they really feel this in another question," he says.

Which brings us right back to the fundamental difficulty with the subject of consciousness. As humans, we understand subjective terms such as 'the self' and what it is to have 'a feeling', but there is as yet no direct relation between these terms and the objective ones used in science.

Onwards and upwards

It would be unfair to say that the conscious robot project failed. Having built the infrastructure, Prof Holland is continuing the work on imagination with another PhD student, Hugo Marques, and is keen to get some additional support. And Gamez has made some headway in predicting the likelihood of a machine being conscious.

In engineering terms, the team have not come up against any brick walls, according to Prof Holland, but there is still further work to do. "When we get the system running reliably with CRONOS moving around freely and doing things in the real world, and SIMNOS running 'inside' CRONOS thinking about alternative strategies, then we'll have to do another mountain of engineering work on developing the structures that compose SIMNOS' internal ideas of itself," he says.

"I think memory, especially episodic memory, will turn out to be particularly important."

An unanticipated outcome of the project has been the success of CRONOS. "It's like sending a rocket to the Moon and discovering Teflon for non-stick frying pans," jokes Prof Holland.

Rob Knight, who built CRONOS, has set up a robotics company in France. And a clone of CRONOS is already taking shape at the Technical University of Munich.

Prof Holland will be leading the ECCEROBOT project and the partners are Technical University of Munich, University of Zurich, The Robot Studio and the Electrotechnical Faculty of the University of Belgrade, Serbia.

The team will be using the same physics-based internal modelling strategy again, along with many of the software architectural components, but this time concentrating mainly on the precise prediction and control of movement rather than on the cognitive aspects of consciousness.

At the end of the project Holland intends to use the improved modelling technology to take the machine consciousness ideas further.

Comments

All comments

You need to be registered with the IET to leave a comment. Please log in or register as a new user.