STUART ROBINSON/UNIV. SUSSEX

Q&A: Andy Clark

A philosopher's view of robots

Where does the mind stop and the rest of the world begin? When Andy Clark, a philosopher at the University of Sussex, UK, asked this question in the 1990s, it was a world without deep learning or smartphones. As technology has developed, his argument that the boundary between cognition and the environment is porous has deepened. He spoke to Nature *about the state of intelligence research and how a truly intelligent machine needs not only a mind, but also a body.*

What has been the most important advance in cognitive science during your career?
There have been two main advances since I joined the philosophy and cognitive-science community in 1984. The first is the development of artificial neural networks, which are computer systems inspired by the way that neurons interconnect in the brain. Then, in the past decade, a particular theory of how the brain works has emerged that is consistent with that research. We now have an idea of the brain as a probabilistic-prediction device. That, for my money, is the most exciting advance.

Why does artificial intelligence (AI) need more data than brains do to perform the same task?
Our brains have been tailored by millions of years of evolution to help us deal with the kinds of object and structure that we're likely to encounter in the world. AI systems, however, start pretty much from scratch. There's also an architectural difference. A lot of deep-learning systems, which use many layers of artificial neurons to progressively extract features from raw data, do not work in a top-down, prediction-driven way, unlike the brain. They work in a more feed-forward way.

What is the difference between predictive and feed-forward approaches?
A feed-forward approach starts with some input, and works its way forwards, network layer by network layer, to deliver a result. It can be trained using plenty of feedback signals, but once trained, it can only map inputs to ever-deeper representations. This means that it can't benefit from the iterative, context-sensitive checking that is so characteristic of the brain, and is the hallmark of biological intelligence.

Is perception guided entirely by information gleaned from the external environment?
This goes straight to the issue of embodiment. The way in which we perceive the world is actually inflected not just by outside information, but also by information from our bodies and brains. It's when those things get together that we experience the world. One of the tasks that the brain seems to be good at is improving its model of the world just by working on itself. That's something the best AI systems aren't as good at. However, you do see something similar with adversarial networks, in which two networks compete against each other, and with networks that play the same game against themselves, again and again.

Will a general-purpose AI system need a body?
Yes — but the exact meaning of 'body' is negotiable. The system will need the ability to act on and make changes to the world, and to test models. It doesn't have to be a real robot in the physical world. It could have a virtual body in a simulated world. What's important will be for it to have something that corresponds to action and perception. That's how you get on top of causality in the world — and if you're not on top of causality, then you're not really understanding the world, even if you're doing amazing things.

Don't AI systems already test and improve their models?
What we have at the moment is something that is close to the limit of passive, non-embodied approaches to AI. These approaches don't test and improve a general-purpose model of the world through real engagement with that world — instead, they are restricted to special-purpose models for a particular domain. A system that is excellent at playing chess isn't the same as the system you would want to pilot your autonomous car. To develop a truly general intelligence, a system has to be capable of embodied interaction with the world.

Will AI systems with a physical embodiment have an advantage over virtual ones?
US philosopher Hubert Dreyfus liked to point out that the physical world has an endless depth. When you look at a wall, you might think, "What's under that brick?" Then you could lift up the brick, see the ants beneath it, and poke at the earth. You could do that for any aspect of the world around you. The simulated world has to bottom out somewhere. So it's still worth putting real robots out into the physical world. But because the mechanical, material challenges of doing so are great, I don't think that will be happening any time soon.

Will culture become necessary for AI systems?
Humans build both physical and cultural systems, and those systems build us. That process has been going on for generations. There is nothing similar for AI systems, yet. Their development will take off when something similar to culture exists for them — some way for them to create the conditions under which they can learn. My best guess would be that we will start to see the most powerful forms of AI emerge when simulated AI agents are able talk to each other as part of proper communities. ■

INTERVIEW BY MICHAEL SEGAL

This interview has been edited for length and clarity.

CORRECTION

The Outlook Q&A 'A philosopher's view of robots' (*Nature* **571**, S18; 2019) wrongly affiliated Andy Clark with the University of Edinburgh. He is at the University of Sussex, UK.