Article:
Prescott, T. (2015) Me in the machine. New Scientist, 225 (3013). pp. 36-39. ISSN: 0262-4079

https://doi.org/10.1016/S0262-4079(15)60554-1

The "me" in the machine

What might we discover about being human by giving a robot a sense of self? Tony Prescott is finding out

What is the self? Rene Descartes encapsulated one idea of it in the 1600s when he wrote: "I think, therefore I am". He saw the self as irreducible and constant, the essence of his being, on which his knowledge of everything else was built. Others have very different views. Writing a century later, David Hume argued that there was no "simple and continued" self through which he experienced the world, just the flow of his experience. Hume's proposal resonates with Buddhist teaching concerning the *anattā*, or non-self, which contends that the idea of an unchanging self is both an illusion and at the root of much human unhappiness.

Today, a growing number of philosophers and psychologists subscribe to the school of thought that the self is an illusion. But even if the centuries-old idea of self as essential and unchanging is misleading, there is still much to explain. How you distinguish your body from the rest of the world, for example. Why you experience the world from a specific perspective – typically, somewhere in the middle of your head. How you remember yourself in the past or imagine yourself in the future. How you are able to conceive of the world from another's point-of-view. I believe that science is close to explaining many of these things.

The key is the insight that the self should be considered not as an essence but as a set of processes – a process being a virtual machine running inside a physical one, as when a program runs on a computer. In a similar way, some of the activity in the brain constitutes processes that generate the human self. This fits with Hume's intuition that if you stop thinking the self vanishes along with the content of your thoughts. So, for instance, when you fall asleep, "you", as an entity brought into being by a set of active brain processes, cease to exist. However, when you awake, those same processes are rekindled, picking up much where they left off, and providing a subjective experience of continuity.

The idea that the self emerges from a set of processes is also what has encouraged my colleagues and I to believe we can recreate it in a robot. By deconstructing the self and then attempting to build it up again piece-by-piece, we are learning more about what selfhood is. This is an on-going collaboration with researchers in several European institutes, and admittedly we still have a way to go. But I'm confident we can create an artificial self – or at least as much of a self as we would like in a robot. We believe our work will help resolve the mystery at the heart of self – that it feels both compellingly real and yet, when examined closely, seems to dissolve away.

Meet iCub, a state-of-the-art humanoid, the robot in which we are creating this sense of self. iCub has vision, hearing, touch and a proprioceptive sense that allows it to coordinate its 53 joints. It can speak and interact with its world, and it improves its performance by learning. There are currently 30 such robots in research labs around the world. At Sheffield Robotics our iCub has a control system modeled on the brain, so that it "thinks" in ways similar to you and me.

Starting from the premise that the self consists of a collection of processes, we first had to consider exactly how we might deconstruct selfhood in order to build it up in a machine. Philosophy, psychology and neuroscience have provided many insights into what constitutes the human self, and how to recognise and measure aspects of self in adults, infants, even animals. Our attempt to construct a robot self begins with psychology, but we will see that it can also be mapped onto our growing understanding of how the psychological self emerges from the brain.

William James, a founding father of modern psychology, suggested that the self can be divided into "I" and "me" – the former comprising the experience of being a self, the latter the set of ideas you have about your self, or the content that is experienced by the "I". In the 1990s, Ulric Neisser, the father of modern cognitive psychology, went further. Beginning with different categories of self-knowledge, he identified five key aspects of self: the ecological or physically situated self, the interpersonal self, the temporally extended self, the conceptual self and the private self. Neisser's analysis is not the final word, but it provides a perspective that is grounded in our understanding of human cognitive development, that was missing in classical philosophical views of the self such as those of Hume and Descartes. It has also provided some useful clues about what might be required to build up an artificial self process-by-process.

How have we gone about creating these self processes for our robot? We use an approach called neurorobotics, which means we incorporate knowledge about how real brains work into our programing. So our iCub's control system is designed to emulate key processes found in the mammalian brain. And the interactions between these simulated brain processes are governed by an architecture called "Distributed Adaptive Control" developed by my colleague Paul Verschure at the Catalan Institute of Advanced Research in Barcelona, Spain, which is modelled on the cognitive architecture found in the brain.

Now, say we want to start building a process that emulates the human ecological self. A key aspect of this is an awareness of ones body and how it interacts with the world. To achieve this, iCub needs an internal "body schema" – a process that maintains a model of its physical parts and the geometry of its current body pose. Rather than programing the body schema directly, as other roboticists might do, we have given iCub the capacity to work it out. Its programing allows it to learn

by generating small, random movements and observing the consequences these have with its various sensors. A similar kind of exploratory behaviour — termed "motor babbling" — is seen in human babies, both in the womb and in early infancy, suggesting that people learn about their bodies in much the same way.

Using this approach, Giorgio Metta and colleagues at the Italian Institute of Technology in Genoa, Italy, are training our iCub to distinguish self from other, a fundamental aspect of the ecological self. The motor babbling program also allows the robot to learn how to move to achieve a specific target pose. Combining this body model with knowledge of objects and surfaces in nearby space enables iCub to move around without colliding into things.

Then there's the temporally extended self. Insights about what this entails can be found in the case of a young man we will know as N.N. who, as the result of an accident in the 1980s, lost his long-term memory for new experiences. This damage to his brain also left him completely without foresight. He described trying to imagine his future as "like swimming in the middle of a lake. There is nothing there to hold you up or do anything with." In losing his past, N.N. had also lost his future. His ecological self remained intact, but we could say that it had become "marooned in the present".

The ability to think about the self in time also poses a problem for our robot. Although we can channel all its sensory input into a large capacity hard-drive, iCub also needs to be able to decide how best to use this information when trying to make sense of the present. Peter Dominey and his group at Inserm in Lyon have addressed this problem by encoding iCub's interactions with objects and people in a way that allows it to more clearly see their relevance to present situations. However, their model uses standard computing techniques, so we are now working with them to create a neurorobotic version. This will directly emulate processing in brain areas, such as the hippocampus, that are known to be important for human autobiographical memory function.

Recent brain imaging studies have confirmed N.N.'s experience that the same brain systems underlie both our ability to recall past events and to imagine what the future might bring. Our hope is that a model of the temporal self will provide iCub with contextual information from the past that will help it to better understand its current experience. This, in turn, should allow it to better predict what could happen next.

Thinking about self as a set of processes, it becomes clear that some of these processes are connected. For example, a key aspect of the interpersonal self is empathy. Empathy derives from a general ability to imagine oneself in another's shoes. One way to do this is to internally simulate what you perceive to be their situation by using the internal model that underlies your own ecological self. So

the interpersonal self could grow out of the ecological self. But what more is needed? We are working on what we consider to be an important building block, the capacity to learn by imitation.

Your ability to interpret another person's actions using your own body schema is partly down to mirror neurons – cells in your brain that fire both when you perform a given movement and when you see someone else perform it. Using this insight, Yiannis Demiris at Imperial College London has extended iCub's motor babbling program into an imitation learning system. As a result, iCub can rapidly acquire new hand gestures, and learn sequences of actions involved in playing games or solving puzzles, simply by watching people perform these tasks. To achieve empathy will require extending the system further so that iCub recognises and mirrors both the physical (movement) and emotional state of the person being observed.

There is still plenty to do. Our models of the ecological, interpersonal and temporal selves are undoubtedly crude in comparison to what goes in human brains. And we have yet to tackle the conceptual and private selves that would provide iCub with knowledge of what (or who) it is, and an awareness that it has an internal world not shared by others. As and when we meet the challenge to make iCub's self processes more realistic, there may be some aspects of the human self that we will not want to emulate. For example, the robot's motivations and goals are essentially those we design in, and it might be wise to leave things that way rather than allow them to evolve over time as they do in people.

Something else holding us back is iCub's limited understanding of language. Although our robot can recognise speech this is not the same as understanding meaning, which requires relating words to action and objects. Our colleagues in Lyon are working on a neurorobotic solution to this problem but for the moment iCub is only capable of two-way conversations on a few topics, such as the game that it is currently playing with you.

That said, even now, we can see the practical potential of robots of this kind. The ecological self makes our iCub safer to be around. The temporally extended self allows it to remember the past and so be better prepared for the future. The interpersonal self allows it to conceive of, and anticipate, human needs and actions. Such a robot could usefully work alongside people in many fields from manufacturing and search-and-rescue to helping care for people with disabilities.

You might argue that in considering what is required to create a sense of self we have missed a crucial element, the "I" at centre of William James' notion of self—what we also call consciousness. But one possibility is that consciousness happens when the other aspects of self are brought together, in other words, it

may be an emergent property of a suitably configured set of self processes, rather than a distinct thing in itself. Returning to the Buddhist idea of the self as an illusion, when you strip away the different component processes that make up the self perhaps there will be nothing left.

Our idea of the self is intimately tied-up with our notion of what it means to be a person. Is it conceivable, then, that one day we might attribute personhood to a robot with an artificial sense of self? Philosopher John Locke, writing in the 17th century, defined a person as an entity with reason and language, possessing mental states such as beliefs, desires and intentions, capable of relationships, and morally responsible for its actions. Modern philosopher Daniel Dennett at Tufts University in Boston largely agrees, but with an important addition. A person, he says, is someone who is treated as a person by others. So, personhood is something we grant to one another.

Note that neither Locke nor Dennett require that a person to be made of biological stuff. This leaves open the possibility that there might, one day, be an artificial entity to which we attribute personhood. Even so, at this stage our iCub falls short of the required criteria. iCub's artificial self can reason, use language, have beliefs, intentions, and enter into relationships of a kind. We might even be inclined to judge it for the appropriateness of its actions. However, our iCub does not yet have the full set of processes associated with a human self, so we cannot be sure that its mental states are anything like ours. Neither is it a moral being – at least not as we commonly think of them – because it does not base its choices on values. On the other hand, our everyday attribution of personhood is grounded more in direct impressions than on a philosophical checklist. As Dennett says, personhood is partly in the eyes of the beholder. And, when interacting with iCub, it can feel natural to behave towards this robot as though we are taking the first steps towards creating a new kind of person. Sometimes it even leaves me with the surprising feeling that "someone is home".

Tony Prescott is Professor of Cognitive Neuroscience at the University of Sheffield, UK, and director of Sheffield Robotics