THERE are a lot of hard problems in the world, but only one of them gets to call itself “the hard problem”. And that is the problem of consciousness – how a kilogram or so of nerve cells conjures up the seamless kaleidoscope of sensations, thoughts, memories and emotions that occupy every waking moment.

The intractability of this problem prompted British psychologist Stuart Sutherland’s notorious 1989 observation: “Consciousness is a fascinating but elusive phenomenon... Nothing worth reading has been written on it.”

The hard problem remains unsolved. Yet neuroscientists have still made incredible progress understanding consciousness, from the reasons it exists to the problems we have when it doesn’t work properly.

Is consciousness still fascinating? Yes. Elusive? Absolutely. But Sutherland’s final point no longer stands. Read on...
We're no longer forced to speculate about the mind, we can watch it in action, says Daniel Bor

"IN THEORY WE COULD CALCULATE HOW CONSCIOUS ANYTHING IS, BE IT A HUMAN, RAT OR COMPUTER"

The first time I saw my father in hospital after his stroke, I was disturbed to find that my strong and confident dad had been replaced by a confused and childlike. Besides being concerned about whether or not he would recover, I was struck by the profound metaphysical implications of what had just happened.

At the time I was a few weeks away from my final university exams in philosophy and neuroscience, both of which addressed consciousness. In my philosophy lectures I had heard elegant arguments that consciousness is not a physical phenomenon and must be somehow independent of our material, corporeal brains. This idea, most famously articulated by Descartes as dualism, nearly 400 years ago, seemed in stark contrast to the neuroscientific evidence in front of me: my father's consciousness had been maimed by a stroke. This idea, most famously articulated by Descartes as dualism, nearly 400 years ago, seemed in stark contrast to the neuroscientific evidence in front of me: my father's consciousness had been maimed by a stroke.

Soon after, I abandoned plans for a PhD in the philosophy of the mind, opting for one on the neuroscience of consciousness instead. There are certainly questions about our minds that seem more in the realms of philosophy. What is it like to be someone confused and childlike. Besides being concerned about whether or not he would recover, I was struck by the profound metaphysical implications of what had just happened.

Seeing red

Those kinds of investigations have been invaluable for narrowing down the search for the parts of the brain involved in consciousness by visual stimuli – indeed, any sensory stimuli. How can we get round this problem? One solution is to use stimuli that are at the threshold of awareness, so they are only sometimes perceived – playing a faint burst of noise, for instance, or flashing a word on a screen almost too briefly to be noticed. If the person does not consciously notice the word flashing up, the only part of the brain that is activated is that which is directly connected to the sense organs concerned, in this case the visual cortex. But if the subject becomes aware of the words or sounds, another set of areas kick into action. These are the lateral prefrontal cortex and the posterior parietal cortex, another region heavily involved in complex, high-level thought, this time at the top of the brain, to the rear. Satisfyingly, while many animals have a thalamus, the two cortical brain areas implicated in consciousness are nothing like as large and well developed in other species as they are in humans. This fits with the common intuition that, while there may be a spectrum of consciousness across the animal kingdom, there is something very special about our own form of it.

In humans the three brain areas implicated in consciousness – the thalamus, lateral prefrontal cortex and posterior parietal cortex – share a distinctive feature: they have more connections to each other and to elsewhere in the brain, than any other region. With such dense connections, these three regions are best placed to receive, combine and analyse information from the rest of the brain. Many neuroscientists suspect that it is this drawing together of information that is a hallmark of consciousness. When I talk to a friend in the pub, for instance, I don’t experience him as a series of disjointed features, but as a unified whole, combining his appearance with the sound of his voice, knowledge of his name, favourite beer and so on – all amalgamated into a single person-object.

How does the brain knit together all these disparate strands of information from a variety of brain locations? This is one of the most prominent attempts to turn this experimental data into a theory of consciousness is known as the “global neuronal workspace” model. This model links consciousness with the fact that there is high activity in our lateral prefrontal and posterior parietal cortices when we carry out new or complex tasks, while activity in these areas dips when we do repetitive tasks on autopilot, like driving a familiar route.

The main rival to global workspace as a theory of consciousness is a mathematical model called the “information integration theory”, which says consciousness is simply combining data together so that it is more than the sum of its parts. This idea is said to explain why my experience of meeting a friend in the pub, with all senses and knowledge about him wrapped together, feels so much more complex, high-level thought, this time at the top of the brain, to the rear. Satisfyingly, while many animals have a thalamus, the two cortical brain areas implicated in consciousness are nothing like as large and well developed in other species as they are in humans. This fits with the common intuition that, while there may be a spectrum of consciousness across the animal kingdom, there is something very special about our own form of it.

In humans the three brain areas implicated in consciousness – the thalamus, lateral prefrontal cortex and posterior parietal cortex – share a distinctive feature: they have more connections to each other and to elsewhere in the brain, than any other region. With such dense connections, these three regions are best placed to receive, combine and analyse information from the rest of the brain. Many neuroscientists suspect that it is this drawing together of information that is a hallmark of consciousness. When I talk to a friend in the pub, for instance, I don’t experience him as a series of disjointed features, but as a unified whole, combining his appearance with the sound of his voice, knowledge of his name, favourite beer and so on – all amalgamated into a single person-object.

How does the brain knit together all these disparate strands of information from a variety of brain locations? This is one of the most prominent attempts to turn this experimental data into a theory of consciousness is known as the “global neuronal workspace” model. This model links consciousness with the fact that there is high activity in our lateral prefrontal and posterior parietal cortices when we carry out new or complex tasks, while activity in these areas dips when we do repetitive tasks on autopilot, like driving a familiar route.

The main rival to global workspace as a theory of consciousness is a mathematical model called the “information integration theory”, which says consciousness is simply combining data together so that it is more than the sum of its parts. This idea is said to explain why my experience of meeting a friend in the pub, with all senses and knowledge about him wrapped together, feels so much more...
than the raw sensory information that makes it up.

But the model could be applied equally well to the internet as to a human. Its creators make the audacious claim that we should be able to calculate how conscious any particular information-processing network is—be it in the brain of a human, rat or computer. All we need to know is the network’s structure, in particular how many nodes it contains and how they are connected together.

**Fiendishly hard**

Unfortunately the maths involves so many fiendish calculations, which grow exponentially as the number of nodes increases, that our most advanced supercomputers could not perform them in a realistic time frame even for a simple nematode worm with about 300 neurons. The sums may well be simplified in future, however, to make them more practical.

This mathematical theory may seem very different from global neuronal workspace—it ignores the brain’s anatomy, for a start—and yet encouragingly both models say consciousness is about combining information, and both focus on the most densely connected parts of the information-processing network. I feel this common ground reflects the significant progress the field is making.

We may not yet have solved the so-called hard problem of consciousness—how a bunch of neurones can generate the experience of seeing the colour red. Yet to me, worrying about the hard problem is just another version of dualism—seeing consciousness as something that is so mysterious it cannot be explained by studying the brain scientifically.

Every time in history we thought there had to be some supernatural cause for a mysterious phenomenon—such as mental illness or even the rising of bread dough—we eventually found the scientific explanation. It seems plausible to me that if we continue to chip away at the “easy problems” we will eventually find there is no hard problem left at all.

**ALTERED STATES**

Sometimes strange states of consciousness can reveal a lot about how we normally construct reality.

**Liar, liar**

Why did you pick that outfit to wear this morning? What made you do your to-do list in that order today? In fact, how did you even end up in that job? You may think you know the reasons, but they could be a work of fiction.

That bizarre conclusion has emerged from studies of people who have had an extreme form of brain surgery—the complete severing of the thick bundle of nerves connecting the two hemispheres of the brain—in an attempt to cure their epilepsy.

Such people usually seem fine, but tasks that test cross-talk between the hemispheres can catch them out. In one test, people had different images shown to each eye, and had to point to a similar image with the hand on the same side as the eye.

When one person saw a snow scene with his left eye, he chose a picture of a snow shovel with his left hand. But when asked to explain his choice, he had a problem. His left eye and hand’s actions were under the control of his right brain, as each brain hemisphere controls the opposite side of the body. But language is controlled by the left brain, which could not access the snowy image “seen” by the right brain. So the subject invented a reason that had nothing to do with snow: the shovel was for cleaning out a chicken coop, he said, as a chicken was the last image seen by his left brain.

Such findings have led to the “interpreter theory,” which says that the brain makes up narratives about our actions to help us make sense of the world. Any act or event can be tipped into this sort of confusion. In one study, people who have never had brain surgery were told to choose a picture from a selection, then tricked into thinking they had picked another. When asked for their reasoning, their explanations were convincing—and yet had to be entirely imaginary. Who could say in a split second how often our consciousness plays these sorts of tricks on us? 

**Being conscious that we are conscious seems to be what sets us apart from other animals, as Emma Young discovers.**

**Blindsight**

Crucially, the contrast of the stripes was adjusted for each person so that, no matter how good their vision, everyone got about 70 per cent of the answers right. That meant that for the confidence ratings, the only variable was people’s metacognitive abilities, giving the first demonstration in the lab that this ability varies widely (Science, vol 331, p 1541).

As well as doing these tests, the volunteers also had their brain scanned, and this revealed that those with the best metacognitive abilities had more grey matter in an area in the front of the prefrontal cortex, known as the rostral prefrontal cortex (RFC). This lies just behind the forehead.

“What is it about this region that gives us this ability?” asks Fleming. “Could the fact that it is more developed in humans mean that we have a fundamentally different self-awareness to animals?”

The other classic way of understanding how the brain works is to look at what happens when it is not functioning as it should. Take “blindsight”, a very rare condition usually caused by brain damage.
Injury. Those affected act as though they are, to all intents and purposes, sightless. But careful testing reveals they can take in some visual information about the world at an unconscious level. When asked to guess what object is in front of them, for instance, they do better than if they had just guessed randomly—insisting all the while that they can see nothing.

Blindsight has always been thought to arise from damage to the visual cortex, at the back of the brain, where information from the optic nerves first arrives. Recent brain imaging studies, however, suggest that the damage also affects connections to the prefrontal cortex, the same region that brain imaging studies, however, suggest arises from damage to the visual cortex, at the back of the brain, where information from the optic nerves first arrives. Recent brain imaging studies, however, suggest that the damage also affects connections to the prefrontal cortex, the same region highlighted by Fleming. “That’s a pretty big rethink,” says Hakwan Lau at Columbia University in New York City, who led the work.

Such a change of heart would sit better with Fleming’s work, as blindsight is one of the starkest failures of metacognition it is possible to imagine.

### Out of control
Less extreme impairments of metacognition may be involved in other common disorders, such as schizophrenia, which involves delusions and hallucinations. “Some individuals have a problem with that very central metacognition, that I know I’m me and I know what I’m doing,” says Janet Metcalfe, also at Columbia University. She has studied the metacognitive abilities of people with schizophrenia using a simple cursor-based computer game at first the schizophrenia patients were as good at judging how well they performed as the group of healthy control subjects. But when Metcalfe started secretly moving the cursor herself, the control group quickly recognized something strange was going on. People in the schizophrenia group, on the other hand, failed to realize that they were no longer completely responsible for the cursor’s movements.

Some people with schizophrenia come to believe that others are controlling their behaviour and have a belief in a microchip has been implanted inside their head. “If you don’t know you’re controlling your own behaviour, you could be open to that kind of symptoms,” says Metcalfe.

As well as potentially helping people with schizophrenia, better understanding of metacognition may be involved in teaching in schools. Metcalfe has found that children aged from about 7 to 11 are capable of making good metacognitive judgements about how well they know a subject, but may fail to make use of that knowledge. For instance, such children prefer to continue spending time on a subject they already know rather than moving on. “It would be interesting to try to teach teachers how to use metacognition effectively,” says Metcalfe.

It may be possible to improve people’s metacognitive abilities, by giving them feedback after the kind of computer tasks used by Fleming. Metcalfe hopes this will help people with schizophrenia. But suppose the rest of us did the same kind of training. Would that give us a supercharged super-consciousness?

“If you define consciousness as what it’s like to experience a train or a journey, then let’s not go on changing that,” says Fleming. “If it’s being able to accurately reflect on what you’re seeing, or whether you just made a good decision,” says Fleming, “then training could give it a boost.”

Emma Young is a writer based in Sheffield, UK.

---

### The twilight zone

Fancy experiencing an altered state of consciousness without getting high on hallucinogenic drug? Easy. We slip into such a state every night when we sleep. To know how to appreciate it at the time becomes, well, unconscious.

This set-up revealed that neuronal activity preceding people’s conscious decision to press the button by nearly half a second. More recently a similar experiment placed people in an MRI scanner instead of hooking them up to electrodes. This found stirrings in the brain’s prefrontal cortex up to 2 seconds before someone became aware of having made a decision (PloS One, vol 6, p e16216).

These results are sometimes interpreted as disproving the existence of free will. They could equally well mean that we do have free will, but that it is our unconscious mind that is, in fact, in charge, not the conscious one. Neuroscientist John-Dylan Haynes at the Max Planck Institute in Leipzig, Germany, who led the brain scanning study, warns against jumping to this conclusion. “I wouldn’t interpret these early [brain] signals as an ‘unconscious decision,’” he says. “I would think of it more as an unconscious bias of a later decision.”

Unfortunately, it is tricky to analyse mental processes that are outside our conscious awareness. Some researchers have even resorted to using on-line tasks to try to communicate with people’s unconscious. And they have had some success, too.

A more orthodox technique is to use a method called “masking,” in which an image is flashed in front of the eyes only to be quickly replaced with another before the unconscious has a chance to register. In this way it has been demonstrated that information shown to the unconscious is then taken into conscious thoughts and decisions. For instance, people shown the masked word “salt” afterwards pick a saltshaker, while a related word from a list, like “pepper.”

Asking people to choose words from a list might seem a rather artificial test, but such unconscious associations can spill over into life outside the research lab. One study of 400 women, for instance, that people acted more competitively in a game if the pen and paper were taken out of a briefcase, but not if it was stuck to a backpack. Afterwards no one was aware of it affecting their behaviour.

### Make your mind up time

If these ideas are disconcerting, they may be an upside, points out Ap Dijksterhuis of Radboud University Nijmegen in the Netherlands. Our ability to unconsciously process information may help us to make decisions.

The idea was tested by Dijksterhuis, for instance, people were asked to choose an apartment by one of three methods. They were asked, for instance, to imagine a briefcase, to make an instant decision, or to evaluate all the pros and cons for a few minutes, or thinking about an apartment while walking. Dijksterhuis then asked them to inconspicuously fill out a form on the number of times they would consider that apartment in the future. Dijksterhuis reckons that unconscious deliberation can also explain “aha” moments when the answer to a problem seems to come from nowhere, as well as times when a searching for words comes to mind only after we stop trying. “For all sorts of decisions we are never aware of all the myriad of influencing factors,” says Haynes.

Caroline Williams is a writer based in Sunday, UK.
So much time is spent working out what consciousness is, says Clare Wilson, that we sometimes neglect an equally important question...

**What is consciousness?**

Despite this catch 22, we have made some inroads into discovering what consciousness is, at least in terms of what we can see going on in the brain using MRI scanners and electrodes on the scalp. One of the leading theories, the global neuronal workspace model, says that sensory stimuli, such as sights and sounds, are initially processed separately and locally at an unconscious level. Only the most salient information emerges into consciousness when it ignites activity in broader networks of neurons across the rest of the brain (see page 32).

According to this theory, the function of consciousness is to carry but difficult or complex mental tasks – ones that require information from multiple sources to be combined and integrated. Daniel Bor, a neuroscientist at the Sackler Centre for Consciousness Science in Brighton, UK, thinks this is plausible, but has a slightly different take on it. He believes that a key function of consciousness is to combine information in a way that leads to innovation and problem-solving, in particular through a process known as “chunking”.

We can usually hold only a few things in our working memory at once, but if we chunk together related items it is easier to manipulate more concepts simultaneously. “Maybe consciousness is a way of binding components together in order to chunk them,” says Bor.

The idea is speculative but there is some supporting evidence. Two of the three areas of the brain seen as the home of consciousness – the prefrontal and parietal cortices – light up more strongly for mental tasks that require chunking than for anything else (Frontiers in Psychology, vol 3, p 64).

Another strand of support comes when we consider whether any other creatures are conscious: “in theory all other animals could be zombies. Yet many people who study consciousness do not see it as an all-or-nothing quality: while other animals may not have the highly developed and special form of consciousness that we have, some species probably have a glimmer of it. And those animals we think of as most likely to be conscious – apes and dolphins, for instance – are also innovative problem-solvers and toolmakers. “I think that’s an important clue,” says Bor.

These are not the only theories about the function of consciousness. In the 1970s, the idea emerged that it was the need to understand other people’s minds that made us aware of our own. “It is more difficult to anticipate the perceptions of others if you cannot perceive your own,” says David Barash, a psychologist at the University of Washington in Seattle. That might suggest human consciousness scaled greater heights as our ape-like ancestors started living in larger social groups, with the ensuing daily potential for aggression and competition.

It is not the only theory that relates the evolution of consciousness to group living. But for neuroscientist Chris Frith of University College London, the benefits of cooperation rather than competition. “It’s so that we can talk to each other about experiences,” he says.

**Combined senses**

Frith’s group has shown that people make better decisions in laboratory tasks if they are allowed to mull over the pros and cons of the evidence with a partner. (Science, vol 329, p 5955). That might sound obvious, but it is hard to imagine a zombie being able to do so as it requires reflection and introspection, key traits of consciousness. “We have to be able to reflect upon our experiences before we can talk about them,” says Frith.

Rees, who works with Frith, gives the example of two early humans regarding a distant dust cloud and trying to work out if it signals a herd of buffalo or a pack of lions. The better they are at reflecting on their feelings and judgements, the better their collective decision-making about whether to go to it or flee. You can combine the forces of your sensory systems that becomes a useful adaptive strategy.

Yet Frith thinks a better example of the benefits of consciousness would be the earliest. “We are introduced to the characteristic flavour of buffalo meat, and thereby deducing where the herd had been grazing.”

Of course, consciousness could have evolved for multiple reasons – or perhaps only one. “The idea that consciousness is an automatic property of intelligence,” says Bora. “Yet that can feel like a cop-out. “My guess is that consciousness, because of its complexity and confusability, in fact conferred adaptive value on its possessors,” says Barash, “but I can’t think of any way to prove it.”

---

**39 | NewScientist | 18 May 2013**

**Out of hand**

In the 1960s film Dr Strangelove, the lead character had a bizarre affliction. His right arm seemingly had a mind of its own. Such a condition very discuss, although it vanishingly rare.

People with so-called anarchic hand syndrome find that their affected limb reacts out and grabs things they have no wish to pick up. They might try restraining it with their other hand, and if that doesn’t work, “they sometimes come to the line and kick back and tied up,” says Sergio Della Sala, a neuroscientist at the University of Edinburgh, UK, who studies the condition.

The cause is injury to the brain, usually in a region known as the supplementary motor area (SMA). Work on monkeys has shown that another part of the brain, the premotor cortex, generates some of our actions unconsciously in response to things we see around us. The SMA then kicks in to allow the movement or stop it, but damage to the SMA can wreck this control – hence the anarchic hand, acting on every visual cue.

A few people are unfortunate enough to have damage to the SMA on both sides of the brain, and experience both hands acting outside their control. They are at the mercy of environmental triggers, says Della Sala.

The system works like the very opposite of free will – Della Sala calls it “forced will.” The findings suggest that, while it feels like our actions are always under our conscious control, in fact there is a lot of unconscious decision-making going on too.

If that sounds implausible, have you ever been driving somewhere on a day off and found yourself heading towards the office at the moment you hit part of your normal routine to work? That’s your prefrontal cortex responding to an environmental cue right there.

Clare Wilson

---

Clare Wilson, that we sometimes neglect an equally important question...
We will only understand consciousness once we can give it to machines, says Celeste Biever

FROM C-3PO of Star Wars to Wall-E, the sentient garbage collector, the prevalence of conscious machines in the stories we tell seems to reflect humanity’s deep desire to turn creator and design an artificial intelligence. It might seem as if we stand little chance of making an artificial consciousness when the natural variety remains such an enigma. But in fact, the quest for machine consciousness may be key to solving the mystery of human consciousness, as even scientists outside AI research are starting to acknowledge.

“The best way of understanding something is to try and replicate it,” says psychologist Kevin O’Regan of Descartes University in Paris, France. “So if you want to understand what consciousness is, well make a machine that’s conscious.”

University in Paris, France. “So if you want to understand what consciousness is, well make a machine that’s conscious.”

AI research are starting to acknowledge. The much-hyped Human Brain Project, based in Geneva, Switzerland, aims to try and design an artificial intelligence. But not before it had come to the conclusion that incoming sensory information and other low-level thought processes initially stay in the unconscious. Only when information is salient enough to enter the global workspace, do we become aware of it in the form of a conscious “broadcast” to the whole brain. Since Baars proposed this idea, in 1987, numerous strands of supporting evidence have accumulated, including that derived from scanning the brains of people under anesthesia.

While Baars stumbled on the AI work that informed his theory, some computer scientists are now deliberately trying to copy the human brain. Take a software bot called LIDA, which stands for learning intelligent distribution agent. LIDA has unconscious and conscious software routines working in parallel, designed as a test of global workspace principles. But in this case the term “conscious” does not mean that the program is conscious just that it broadcasts the most important results across all subroutines.

Nitty gritty

The much-hyped Human Brain Project, based in Geneva, Switzerland, aims to build a functioning software simulation of the entire human brain on a 10 cubic millimetre chunk of rat brain, but in February they won a $1 billion grant, which they reckon will take them to simulating a whole human brain.

But copying the brain’s architecture misses the point, O’Regan thinks. “How the brain is organised isn’t the interesting question,” he says. “The interesting question is why do we feel?” In other words, how can electricity moving through neurons create the subjective feeling of pain, or the colour red? “It’s not just that we know we are in pain, there is the real, nitty gritty feel,” he says. Trying to create a machine that experiences pain or colour in the same way that we do might require a radical rethink. Pentti Haikonen, an electrical engineer and philosopher at the University of Illinois in Springfield, believes that we will never create a feeling machine using software. Software is a language, he says, and so requires extra information to be interpreted. If you don’t speak English, the words “pain” or “red”, for instance, are meaningless. But if you see the colour red, that has meaning no matter what your language.

Most computers and robots created so far run on software. Even if they connect to a physical device, like a microphone, the input has to be translated into strings of 1s and 0s before it can be processed. “Numbers do not feel like anything and do not appear as red,” says Haikonen. “That is where everything is lost.” Not so for Haikonen’s robot. His machine, called XCR for experimental cognitive robot, stores and manipulates incoming sensory information, not via software, but through physical objects – in this case wires, resistors and diodes. “Red is red, pain is pain without any interpretation,” says Haikonen. “They are direct experiences to the brain.” XCR has been built so that, if hit with sufficient force, the resulting electrical signal makes it reverse direction – an avoidance response corresponding to pain, Haikonen says. The robot is also capable of a primitive kind of learning. If, when it is hit, it is holding a blue object, say, the signal from its blue-detecting photodiode permanently opens a switch. From now on, the robot associates the colour blue with pain and reverses away. Watch the robot in training and it is hard not to feel sympathy as it is whacked with a stick. “But,” it intones. “Me hurt, blue bad.” The next time Haikonen tries to push the robot towards a blue object, it backs away. “Blue, bad.” The next time Haikonen tries to push the robot towards a blue object, it backs away.

As Haikonen’s robot demonstrates, it is possible to create a machine that is conscious.

How could we ever be sure that an android is conscious?

It’s hard not to feel sympathy as the robot is whacked with a stick. “Me hurt, it intones.”

But is Haikonen’s robot conscious? It might seem as if we stand little chance of making an android conscious, but a recent study at the University of Illinois in Springfield suggests that even machines can feel pain. Haikonen’s robot is capable of short-term memory and the ability to learn. For example, it can learn to associate a blue object with pain. But Haikonen’s robot is not conscious. It is a machine.

As robot achievements grow, learning to avoid a blue object is no big deal: conventional software-based robots can do it standing on their heads. But the fact that XCR bypasses software, storing sensory information directly in its hardware, takes it the first step down the road to awareness, claims Haikonen. “The contents of the consciousness is limited,” he says, “but the phenomenon is there.”

Brain in a vat

It’s a claim that Haikonen makes very cautiously, and one that has not yet convinced many others. “I would hesitate to call something conscious that has such a limited repertoire of responses,” says Murray Shanahan, who studies machine consciousness at Imperial College, London. Still, it’s a new approach, and the first time that such a claim has been made by any serious AI researcher.

If Haikonen is right, and we can’t create a feeling machine based on software, then no matter how big the net gets, it will never be sentient. But a brain in a vat wired up to a supercomputer simulation—a classic thought experiment from philosophy—could be conscious. Haikonen does not say awareness needs a physical body, just a physical brain.

Whether machines of the future run on software or physical brains like those of Haikonen’s devising, how would we know if they do achieve sentience? Self-awareness, by definition, a highly subjective quality. The answer is simple, says O’Regan. Once they behave in the same way we do, we will simply have to assume they are as conscious as we are. If that sounds preposterous, don’t forget, it’s the same assumption we make routinely about our fellow humans every day of our lives.

After all, if you somehow got talking to an alien and you had a similar conversation to one you might have with a person, “you would probably agree that he was conscious—even if it turned out there was cottage cheese inside his brain”, says O’Regan. “The underpinnings of his behaviours are irrelevant.” “When you say that I am conscious, that’s what makes me conscious,” says O’Regan. “It’s hard not to feel sympathy as the robot is whacked with a stick. ‘Me hurt, it intones.’

Celeste Biever is New Scientist’s deputy news editor.