State of the art: Consciousness

Is a bat conscious? **SUSAN BLACKMORE** *argues that there must be something radically wrong with the way we are currently thinking about consciousness, or we would not find ourselves with seemingly intractable problems.*

WHAT is it like to be a bat? This is the question asked by philosopher Thomas Nagel in his famous 1974 paper. As he explains, if there is something it is like to be the bat – something *for* the bat – then the bat is conscious. That is what we mean by consciousness.

Ask 'what is it like to be a stone?' and most people will reply 'nothing'. Ask 'what is it like to be my best friend?' and most people will have a good guess – we believe there is something it is like to be another person. The interesting problems start when we ask about babies, computers or other species. This is why Nagel chose the bat, with its complex world of reflected sounds. The bat's world must be very different from ours; but surely it has subjective experiences, doesn't it? This is the question.

In the 1989 *International Dictionary of Psychology* Stuart Sutherland wrote:

> *Consciousness is a fascinating but elusive phenomenon; it is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written on it.*

In the subsequent decade there was an explosion of research and theorising on consciousness and some of it *is* worth reading. This much has been achieved – people generally now agree that when we talk about consciousness, or the problems of consciousness, we are talking about subjectivity. Work on memory and learning, perception or emotions, may be relevant but is only about consciousness if we are concerned with subjective experiences. Beyond that, the agreement is less secure.

What makes consciousness such a special and intractable problem? The answer is that whichever way you look at it the dualism trap seems to be waiting. If you think that mind is something different from matter then you have problems (as Descartes did) with how the two different worlds interact. If you deny a separate mind and stick to only physical brains and neurons, then you deny your own subjective experience. If you accept that there is subjective experience, and that there are also physical brains that cause that experience then you have to bridge the 'explanatory gap', or what William James called the 'fathomless abyss' or the 'chasm between the inner and the outer worlds' (James, 1890/1983).

The modern version of this old chestnut is what philosopher David Chalmers (1995) calls 'the hard problem', that is, 'how physical processes in the brain give rise to subjective experience'. The 'easy problems', he says, are to do with how the brain works, or the neural underpinnings of perception, emotion or memory. But answers to these (however detailed and accurate they may be) do not explain how my subjective experience of red can *be* the firing of cells in this visual cortex, or how co-ordinated neural oscillations could *cause* my thoughts.

Some people argue that there is no hard problem (O'Hara & Scutt, 1996): once we solve the 'easy' problems consciousness will go the way of caloric fluid or the *elan vital* (Churchland, 1996; see also Shear, 1998). They are probably the majority among those who work on the neural correlates of consciousness or on problems such as the 'binding problem' – that is, how the separately processed features of objects are bound together to make a perceptual whole (Crick, 1994). Others are convinced that the problem is so difficult that it requires a revolution in physics (Penrose, 1994) or even that it is beyond the capacity of our human brains to solve (McGinn, 1999). Perhaps the most helpful conclusion is that there must be something radically wrong with the way we are currently thinking about consciousness or we would not find ourselves with this seemingly intractable problem. A few experiments – and thought experiments – may help.

Blindsight and the zombie

Damage to part of the visual cortex, V1, on one side of the brain causes a blind patch, or scotoma, on the opposite side of the visual field. If an object is presented in this area patients consciously see nothing at all. Lights can be flashed, objects moved, or writing displayed and they will insist that they see nothing. Yet detailed experiments show that, while denying all visual experience, they can nevertheless point to the location of a flashed light, or discriminate at better than chance levels between upward and downward movement, vertical and horizontal stripes, or different objects. This is the odd condition known as 'blindsight' (Weiskrantz, 1986). Perhaps the most common explanation is that the destruction of V1 leaves several other visual pathways intact, and information in these can direct eye movements or other responses that make reasonably accurate guesses possible.

Blindsight has been studied since the early 1970s, provoking numerous experiments and heated debate (Kentridge, 1999) and becoming a popular topic in consciousness studies. The condition seems so weird and counter-intuitive. How can a person respond to something they cannot see? How can they act without having first been aware? Doesn't consciousness have to come first and action depend upon it? These are the very intuitions that are challenged by so much of the modern work on consciousness.

We can sharpen these intuitions by thinking about the philosopher's zombie – an all-time favourite among thought experiments. Imagine you meet me in the

street. I look like Sue Blackmore, I speak like Sue Blackmore, I behave in every conceivable way like a real human being, but I am not conscious. There is no view from within. There is nothing it is like to be me. This (not something gruesome emerging from a Haitian grave) is the philosopher's zombie.

It is easy enough to imagine such a zombie, and to think of the blindsight patient as a partial zombie. But if you think a little harder you will realise that you face two main options. Either you accept that your imagined creature is, in principle, possible. For example, a sophisticated robot might be like this, so might some animals. If you think this way you imply that consciousness is something extra – something additional to the processes that produce behaviour. This leads to all sorts of tangles. You may be tempted to ask why destroying one small part of V1 abolishes consciousness and what is special about the 'consciousness bits' of the brain. More generally, you may ask why, if consciousness is an optional extra, we have it at all. What does it do? Why did evolution give us conscious experience when we might have ended up as zombies instead?

The functionalist's alternative is to say that (imaginable or not) the zombie could never exist. Anything that could behave and speak like Sue Blackmore simply has to be conscious in just the way I am. The blindsight patient does not behave like a normal person. He has only vestigial abilities and a damaged experience to match.

The American philosopher Dan Dennett (1991) has possibly the clearest view on this one. He claims that we are all zombies. We are machines with information-processing brains that produce higher-order representations of our lower-order processes. These allow us to describe ourselves as having thoughts, feelings and so on, and this is what it means to be conscious. Any person or animal or machine that had the appropriate machinery would be conscious in the way that we are, and for the same reasons. This view does away with the tangles, but to some it seems to strip consciousness of its magic. On this view consciousness is no special extra – it just comes with the territory.

The evolution of consciousness

Why did consciousness ever evolve? Why aren't we, and all other animals, just going about our business without any inner experience – without all that suffering, minding and feeling? This is one of those difficult questions that lies right at the heart of the problem of consciousness and has received no satisfactory answer. It is tantamount to asking about the function of consciousness.

If consciousness has a function – if it does something – then natural selection could get to work on it. Conscious creatures would have a selective advantage, and so would pass on their genes for being conscious.

Several theories provide a possible function. For example, Cambridge psychologist Nicholas Humphrey argues that our early hominid ancestors were social creatures who needed to be able to predict each other's behaviour. The best way to do that, he argues, is by observing your own inner processes. So this is why we acquired what he calls the 'inner eye'. Other similar theories relate consciousness to the development of Machiavellian intelligence (in particular the ability to deceive others), and to the development of a 'theory of mind'. Just as children gradually develop the understanding that other people have desires, intentions and points of view, so this ability gradually evolved in the past. Other theories give consciousness the function of alerting the system, reacting to emergencies, or dealing with novelty.

The trouble is that one can still ask why these various abilities needed to be associated with consciousness. Why could our ancestors not predict each other's behaviour, develop a theory of mind, or cope with emergencies unconsciously? Why would subjectivity help them along? Now remember the zombie. If you believe in the philosopher's zombie, then you believe it is possible that we might have evolved the way we have, doing everything the way we do, but without any corresponding consciousness. But this is magic. This makes consciousness some kind of optional extra or epiphenomenon which would be invisible to evolutionary processes.

A convincing theory of the evolution of consciousness has to explain why the evolution of certain abilities simply must mean the appearance of consciousness – rather like the way a chemical theory of H_2O explains when and why water behaves like a liquid and feels wet. Though we do not have a complete functionalist theory of this kind, the implication would be clear. Consciousness *per se* would not have any function because it simply comes with being the kind of organism we are. This would seem to threaten our natural feeling that our consciousness has some kind of power – that consciousness does things. Does it?

Voluntary action and the sense of free will

Hold your arm out in front of you and then, whenever you feel like it – of your own free will – flex your wrist. Do it several times and watch what happens in your own mind. You may feel as though you first consciously decided to move, and then moved. It feels as though the conscious decision caused the action. Trying it for yourself helps with thinking about a famous experiment on voluntary action. Libet (1985) asked people to do just this, while he systematically measured the timing of three things: the start of the action (using electrodes on the wrist), the start of the readiness potential in motor cortex (using electrodes on the scalp), and the decision to move (using a revolving spot on a clock face – subjects had to say where the spot was when they consciously decided to act). The last of these was the most controversial but subsidiary experiments showed that people are able accurately to time external stimuli this way. The assumption is that they could do the same with their own private decisions.

So – which came first? The decision to move, or activity in motor cortex? The answer was more dramatic than anyone expected. The brain activity began about half a second before the person was aware of deciding to act. It seems that the conscious decision came far too late to be the cause of the action: as though consciousness is a mere afterthought. Odd though this might seem, it fits with previous experiments on exposed brains, in which Libet demonstrated that about half a second of continuous activity in sensory cortex is needed for a person to become aware of a sensory stimulus (Libet, 1981). This implies the odd conclusion that consciousness lags behind the events of the world. But, Libet argued, once events reach neuronal adequacy (i.e. half a second of activity) they are subjectively referred back to the time of the initial evoked potential. So even though consciousness takes half a second to build up, events still seem to happen in real time.

Reaction to the study was heated, and still goes on. Numerous commentators argued that the voluntary wrist action is not analogous to free actions in real life, that the clock task is flawed (see commentators on Libet, 1985), or that the whole idea of there being a time at which consciousness

happens is misguided (Dennett, 1991). In some ways, though, no one ought to have been surprised. If you expected consciousness to have started the process then you are really a believer in magic – in some kind of force that acts on brain stuff. Very few people would defend such a view and yet they are surprised when shown, in this vivid way, that brain events happen first. Perhaps the reaction reveals how very confused are our ideas about free will, based on our powerful feelings of conscious freedom and control.

These feelings do not necessarily imply real freedom and control. Dan Wegner, of the University of Virginia, has done a series of experiments that induce the feeling of free will (Wegner & Wheatley, 1999). In experiments with a modified ouija board, two participants place their fingers on a small board on top of a computer mouse. While music plays, the mouse roams freely over a table covered with little pictures; when the music stops they must stop the mouse on one picture. On some trials the confederate forces the mouse. When the participant is asked how sure she was that she chose the stopping place herself, even on forced trials she is often convinced that she freely willed the choice herself.

In other experiments using tricks with mirrors and screens, participants watched either their own arms, or someone else's arms in place of their own. When a taped instruction told them, for example, to touch their nose or wave their hand, and the correct action followed, they reported a powerful sense that they had willed the action themselves – even when it was someone else's arm. In other words the feeling of acting freely is not a reliable guide to the cause of an action.

Wegner argues that free will is an illusion created in three steps. First, we are ignorant about how our brains plan actions and carry them out. Second, we become aware of the results of the planning and call these intentions. Finally, the action occurs after the intention and so we leap – erroneously – to the conclusion that 'our' intention caused the action. Not everyone agrees with this interpretation, and there is currently vigorous debate about how neuroscience can contribute to our understanding of free will (Libet *et al*., 1999). Nevertheless, these results should make us question any conclusions we base on our undoubted feelings of conscious control.

Blindness to change

We might have to question far more than that. The phenomenon of 'change blindness' calls into question the very world we think we experience.

I am looking out of the window. I see trees, flowers and the houses over the road. I have the impression that I can take in, and am aware of, most of this scene at once, and that if anything changed I would notice. Because I am a psychologist I also know that only the fovea has detailed vision, but still I assume that somewhere in my brain there is a detailed representation of the whole scene. This rich mental model is my experience, and its changes are – to use William James's phrase – my stream of consciousness. Experiments in 'change blindness' show that this must be wrong.

Early experiments with an eye tracker revealed a very strange effect. People were asked to read some text and then, when the eye tracker detected an eye movement, the text was changed. Although an observer would see the text flickering and changing all over the place, the reader saw nothing amiss – as long as the letters changed while his eyes were moving. Similar experiments used pictures, and showed that even large changes, such as the movement or disappearance of objects, go unnoticed. From these beginnings the whole topic of 'change blindness' has sprung.

Normally our attention is automatically drawn to a small movement or a single change in a static picture. But anything that masks such changes, or swamps the mechanism, can reveal change blindness, so an expensive eye tracker is not necessary. The whole picture can be moved (Blackmore *et al.*, 1995), or changes can be made during blinks or movie cuts, or brief blank fields can be placed between alternating displays of a scene and a modified version. In all cases a search for the change can take many seconds or even minutes (Rensink *et al.*, 1997). If you think such peculiar effects can only work in the lab, psychologist Dan Simons shows otherwise (Simons & Levin, 1998). Stooges got into conversation with people and then, using clever choreography or distraction, were swapped for a completely different person. About half the time the person talking to them did not notice the substitution.

These results suggest that we never do form detailed representations of the world. Rensink (2000) suggests that focused attention produces a stable representation of just one object at a time, whenever it is needed, making it appear to higher levels as though all the objects in a scene are represented in detail simultaneously, when they are not. We are misled because, if in doubt, we can always look again – using the world itself as an outside memory (O'Regan, 1992). This creates what some have called the 'grand illusion'.

So just how deep is this illusion? Dennett (1991) suggests that the fundamental error is to believe in what he calls 'the Cartesian Theatre'. Theatre metaphors are common in discussions of consciousness, and arguably can be helpful (Baars, 1997). It certainly feels as though I am sitting inside my head and experiencing the events in turn as though they were some kind of show. But this is a big mistake, argues Dennett. While almost everyone rejects outright Cartesian dualism, most psychologists and neuroscientists still believe in some kind of centre, where everything comes together and 'consciousness happens'; some kind of magic finishing line beyond which events 'come into' consciousness; or a centre from where 'my' decisions are made and 'my' instructions sent out. But this cannot be, for the reality of the brain is a massively parallel system with no middle. So, as Dennett (1991) puts it:

> *When you discard Cartesian dualism, you really must discard the show that would have gone on in the Cartesian Theater, and the audience as well, for neither the show nor the audience is to be found in the brain, and the brain is the only real place there is to look for them.* (p.134)

The self

No audience? Is there really no persistent 'me' who lives this life; who is conscious and who has free will? Consideration of the nature of self is deeply bound up with questions about consciousness, as recent debates reveal (Gallagher & Shear, 1999). Philosopher Derek Parfit (1987) divides

WEBLINKS

Center for Consciousness Studies at the University of Tucson, Arizona: *www.consciousness.arizona.edu/*
Information on consciousness by David Chalmers: *www.u.arizona.edu/~chalmers/online.html*
Association for the Scientific Study of Consciousness: *http://assc.caltech.edu.*
Change blindness: *www.wjh.harvard.edu/~viscog/change*

theories of the self into two types – ego and bundle theories. Ego theorists (perhaps the natural way to think) believe in a persistent self who is the subject of experiences and whose existence explains the sense of unity and continuity of experience. Bundle theorists (named after Hume's (1739–40/1986) 'bundle of sensations'), deny there is any such thing. The apparent unity is just a collection of ever-changing experiences tied together by such relationships as a physical body and memory. While ego theories come easily to most of us, intellectually some kind of bundle theory seems ever more inescapable.

What will this mean? Unlike most areas of psychology this argument cannot remain entirely intellectual. When we start to question the very nature of our selves we inevitably start to change ourselves. When we play with theories of consciousness, consciousness itself starts to alter. Methods such as meditation and mindfulness have long been said to break down the false idea of a persistent self, and now some psychologists are beginning to bridge the gap between spiritual practice and academic psychology (Austin, 1998; Blackmore, 1999; Claxton, 1996). Mystics and sages for millennia have described experience without self, the transcendence of self, or have taught the view that the ordinary self is an illusion – and the ultimate cause of suffering. Perhaps our research is leading to the same daunting conclusion.

The future of consciousness

Behaviourism abolished all talk of consciousness from psychology for half a century, but consciousness is back and unlikely to go away again. So where does the future lie? Neuroscience is developing fast. Artificial intelligence and neural modelling are increasing our understanding of the brain. Improvements in brain scanning provide ever more detail of the neural correlates of consciousness. But is this all just working on the 'easy problems'? Will the really hard problem remain to haunt us forever?

I doubt it. I think that one day psychologists will look back and laugh at the silly muddle we got ourselves into. To them the way out will be obvious. The trouble is that right now, like everyone else in the field, I cannot see it.

■ *Dr Susan Blackmore is at the Department of Psychology, University of the West of England, Bristol BS16 2JP.*

References

Austin, J. H. (1998). *Zen and the brain.* Cambridge, MA: MIT Press.

Baars, B. J. (1997). *In the theatre of consciousness: The workspace of the mind.* Oxford, OUP.

Blackmore, S. (1999). *The meme machine.* Oxford: Oxford University Press.

Blackmore, S. J., Brelstaff, G., Nelson, K., & Troscianko, T. (1995). Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception, 24*, 1075–1081.

Chalmers, D. J. (1995, December). The puzzle of conscious experience. *Scientific American*, 62–68.

Churchland, P. S. (1996). The hornswoggle problem. *Journal of Consciousness Studies, 3*, 402–408. (Also reprinted in Shear, 1998)

Claxton, G. (Ed.). (1996). *Beyond therapy: The impact of Eastern religions on psychological theory and practice.* Sturminster Newton, Dorset: Prism Press.

Crick, F. (1994). *The astonishing hypothesis.* New York: Scribner's.

Dennett, D. C. (1991). *Consciousness explained.* London: Little, Brown & Co.

Gallagher, S & Shear, J. (Eds.). (1999). *Models of the self.* Thorverton, Devon: Imprint Academic.

Hume, D. (1986). *A treatise of human nature.* Harmondsworth: Penguin. (Original work published 1739–40)

James, W. (1983). *The principles of psychology.* Cambridge, MA: Harvard University Press. (Original work published 1890)

Kentridge, R. W. (Ed.). (1999). Papers on 'blindsight'. *Journal of Consciousness Studies, 6*, 3–71.

Libet, B. (1981). The experimental evidence of subjective referral of a sensory experience backwards in time. *Philosophy of Science, 48*, 182–197.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8*, 29–539. [See also the many commentaries in the same issue, 539–566, and *BBS, 10,* 318–321]

Libet, B., Freeman, A., & Sutherland, K. (1999). *The volitional brain: Towards a neuroscience of free will.* Thorverton, Devon: Imprint Academic.

McGinn, C. (1999). *The mysterious flame: Conscious minds in a material world.* New York: Basic Books.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83*, 435–450.

O'Hara, K., & Scutt, T. (1996). There is no hard problem of consciousness. *Journal of Consciousness Studies, 3*, 290–302. [Also reprinted in Shear, 1998]

O'Regan, J. K. (1992). Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461–488.

Parfit, D. (1987). Divided minds and the nature of persons. In C. Blakemore & S. Greenfield (Eds.). *Mindwaves* (pp. 19–26). Oxford: Blackwell.

Penrose, R. (1994). *Shadows of the mind.* Oxford: Oxford University Press.

Rensink, R. (2000). The dynamic representation of scenes. *Visual Cognition, 7*, 17–42.

Rensink, R. A., O'Regan, K. O., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368–373.

Shear, J. (Ed.). (1998). *Explaining consciousness – The hard problem.* Cambridge, MA and London: MIT Press.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review, 4*, 644–649.

Wegner, D. M., & Wheatley, T. P. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist, 54*, 480–492.

Weiskrantz, L. (1986). *Blindsight: A case study and implications.* Oxford: Oxford University Press.