

Consciousness on the scientific agenda

The convention that the concept of consciousness is so ill-defined that it lies beyond the reach of serious study seems to be losing ground. But definition is still some way off.

How does the brain generate conscious experience? Twenty years ago, most scientists thought this a question for the philosophers; and most philosophers regarded it as a symptom of some kind (though what kind never became clear) of deep linguistic confusion. In sharp contrast to this picture, there was a large measure of agreement among both scientists and philosophers at a recent symposium* that not only is there a real problem of consciousness but that it is a scientific problem and that the time has come for scientists to tackle it.

Hardly anybody today doubts that consciousness is in some way a product of the brain, a product that is intimately connected with the brain's role in behaviour and the processing of information. Cartesian dualism — the notion that brain-stuff and mind-stuff are essentially separate, though able to communicate with each other — has virtually no contemporary followers. Dualism has not, however, vanished without trace. It has left behind the notion of a 'Cartesian theatre' (located by Descartes himself in the pineal gland, the place where brain-stuff and mind-stuff were supposed to communicate): a privileged site at which brain events enter into conscious awareness, those happening elsewhere remaining in a kind of outer darkness.

This notion came under strong attack at this month's meeting. Instead, it was proposed (M. Kinsbourne, Sargent College, and D. C. Dennett, Tufts University, Massachusetts) that consciousness is a property of whatever pattern of distributed neural firing is (in some still obscure sense) 'dominant' at any particular time. This proposal, however, raises the so-called 'binding' problem: what feature of neuronal activity unifies the ramified patterns of individual neuronal events that take place in millions of separate cell-bodies and in even more millions of synaptic junctions at any given instant? And there is a temporal as well as a spatial aspect to the binding problem: neuronal events take place on a timescale measured in milliseconds, whereas the contents of consciousness have an apparent duration at least two orders of magnitude greater. (Perhaps the only important gap in the coverage of topics at the symposium was a consideration of recent proposals^{1,2} that a solution to the binding problem may unexpectedly be provided by quantum mechanics.)

Also more or less absent from the contemporary scene are behaviourism and its

philosophical equivalent, positivism: the notion that, because only behaviour, not conscious experience, is amenable to direct observation, talk about conscious events has no place in the world of science. There has also been a more radical behaviourist view that holds, on the same grounds, that consciousness does not really exist. (One is tempted to add: it is 'just a figment of the imagination'; but that leads to some rather obvious problems.)

I once asked such a radical behaviourist what, on this view, is the difference between two awake individuals, one of them stone deaf, who are both sitting immobile in a room in which a record-player is playing a Mozart string quartet? His answer: their subsequent verbal behaviour. Mercifully, there were no radical behaviourists at the symposium.

But behaviourism too has not vanished without trace. Its modern offspring is a vigorous form of functionalism, which enshrines as the touchstone of the presence of consciousness the Turing test: if you feed symbols (for example, a string of English words) to a machine (or, more particularly, to a digital computer) and a human being and get symbols back from each of them in reply, and if you cannot distinguish between the machine and the person on the basis of their replies, then, if the human being is conscious, so is the machine. On this view, when the machine correctly uses the symbols for 'red', 'pain', 'itch' and so on, it makes no sense to ask further whether the machine has the sensory experiences ('qualia') that belong to these terms.

When this view is held strongly, it is known in the trade as 'strong AI' (artificial intelligence). A famous attack on strong AI — the 'Chinese room argument' — was made some years ago by John Searle³ (University of California, Berkeley). Essentially, the argument demonstrated that computers have only syntax, not semantics; that consciousness is permeated with semantics (one is typically conscious of things such as 'a red rose situated on a table over there'; indeterminate itches and the like are the rarity); so computers cannot, just by computing transformations of symbol strings, achieve consciousness.

On the whole, Searle seems to have won his argument: at any rate, nobody at the meeting made a serious attempt to refute it. Instead, the current form of functionalism goes beyond the mere digital computer, allowing it now to have limbs and sense organs (that is, the computer becomes a robot).

Moreover, proponents of the contemporary functionalist approach, unlike the earlier radical behaviourists, are willing — indeed, eager — to take account of what is known about events in the real brain that lie between input (stimulus) and output (response).

Dennett, for example, argued that one can already take what is known about the way the brain codes and recodes visual stimuli, put it together with existing data from psychophysics and use the knowledge that results to predict, successfully, new visual phenomena, such as illusory experiences. One can even 'inject' such experiences into the brains of experimental animals, as demonstrated in elegant experiments (W. T. Newsome, Stanford University) in which monkeys responded (behaviourally) to microstimulation of a circuit encoding a particular direction of motion in the same way that they had been trained to respond to an external stimulus having the same directional value processed by the retina and what lies behind it.

What more than this, Dennett asked, do we need? There is no theoretical problem about consciousness, we just have to go on gathering new data, and when we have all the details of what goes on in the brain and how this brain activity interfaces with the environment, that will be in and of itself a complete scientific account of consciousness.

Some, however — probably the majority at the meeting — remain unconvinced that it is going to be so simple. For them, more than just dedicated gathering of experimental data will be needed. What is needed is, rather, a new theory (one comparable with, say, the theory of heat for the relation between the gas flame and the boiling of the kettle) that will render the relations between brain events and conscious experiences 'transparent', to use T. Nagel's (New York University) felicitous term. A series of brute correlations would not suffice. This, after all, is the standard set in all other domains of science, so why not here? This new theory is at present unimaginable, but only in the sense that no-one could have imagined relativity or quantum mechanics before *they* were invented, not because we are dealing with the unknowable or a bad language habit.

Jeffrey Gray

Jeffrey Gray is in the Department of Psychology, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, UK.

1. Penrose, R. *The Emperor's New Mind* (Oxford University Press, 1989).
2. Marshall, I. N. *New Ideas in Psychology* 7, 73–83 (1989).
3. Searle, J. *Behav. Brain Sci.* 3, 417–457 (1980).

*Ciba Foundation Symposium No. 174 on Experimental and Theoretical Studies of Consciousness, London, 7–9 July 1992.