

PAUL M. CHURCHLAND

This ill-defined topic, which has yet to find a governing research paradigm, is perhaps best defined by a series of unanswered questions. What distinguishes our *waking state* from such diverse states as deep sleep, trauma-induced coma, and the unconscious state induced by anaesthetics? Also, what distinguishes the brain's *conscious* representations and activities from the vast majority of its representational and computational activities that never ascend to that special status? Further, what structural, representational, or dynamical features of the brain are *responsible* for the emergence of conscious activity? And finally, what special *functions* does it perform that made conscious brains worthy of natural selection in the first place?

Aspirant theories are plentiful, as are thought-provoking empirical data, but no success has joined them decisively. Herewith, a critical summary of some popular suggestions.

*Special location accounts.* Because consciousness is partly or wholly abolished when certain circumscribed brain areas are damaged, it is tempting to identify neuronal activity in those areas as the basis or embodiment of consciousness. For example, unilateral lesions to the thalamic intralaminar nucleus typically produce in the agent a profound hemineglect on the contralateral side, and bilateral lesions yield permanent coma (Bogen 1995). But for any suggested location, we want an *explanation* of what is special about the activities therein. This lacuna leads theorists to a variety of functional accounts.

*Self-representation accounts.* On these, consciousness is said to occur when the brain perceives, or otherwise represents, some of its own cognitive states and activities. Consciousness is thus a form of *metacognition*. The favoured subjects of such occasional meta-representations are thereby elevated into consciousness (Armstrong 1981, Lycan 1987, Damasio 1999). A standard objection is that this wrongly conflates one special case of consciousness—*awareness-of-self*—with the more general form of consciousness displayed when any awake, alert creature is selectively aware of some feature of its external environment. That phenomenon, presumably, need require no *self-representation* at all. Also, the brain is massively engaged in 'monitoring' its own states at all times, whereas only a very few of those states are ever present to consciousness.

*Self-control accounts.* On these, a brain is conscious to the

extent that it is modulating, manipulating, and steering its own cognitive activities. Consciousness is thus the mark of an agent that is partly *autonomous* in generating not just its own behaviour, but its own *cognition* as well (Churchland 1995, Damasio 1999, Taylor 2001). Objections parallel those just given for meta-perceptual accounts. An awake, alert agent can be conscious solely by virtue of steering its own gross *motor* behaviour. And a live brain engages in widespread self-modulation as a matter of course, whereas only a few of the activities thus modulated are ever present to consciousness. A deliberate hybrid of views 2 and 3 might alleviate the second of these two problems (states at the focus of *both* self-perception and self-modulation might be comparatively rare). But the first problem would remain.

*Competition for executive control accounts.* These re-address the brain's control of the physical body. Consciousness is here portrayed as the solution to a serious problem confronting any system as complex as the brain, an organ with diverse subsystems devoted to monitoring and controlling a wide range of internal and external phenomena. These subsystems, it is said, are all in competition with one another for here-and-now control of how the body's motor and sensory systems are to be deployed. The current contents of anyone's consciousness are always the current representations of whichever brain subsystem has managed to elbow aside or eclipse the clamouring competition. Those representations are distinguished by their having at least temporary executive control—over our speech mechanisms and over the body as a whole—and also, perhaps, by their being made candidates for storage in long-term memory (Baars 1988, Dennett 1991). In the awake state, this competition is never-ending, and so the contents of consciousness are typically ephemeral.

One wants, of course, an account of the *mechanism* of such selective dominance or focal attention. Here the suggestions diverge. The *global workspace* account posits a distinct brain area to which the unconscious subsystems continually submit information, which information enters consciousness only when it is taken up as somehow relevant to the computational activities already under way in that focal workspace. Another version eschews any special area, and posits a process of *shifting coalitions* of spatially distributed neural activity, coalitions of neurons temporarily united by their mutual interaction, perhaps (Baars 1988, Leopold and Logothetis 1999), or by a temporary *synchrony* in their physiological activities, a synchrony that yields them a temporary collective dominance over non-synchronous neurons (Crick and Koch 1990, Singer 2000).

*Special architecture and dynamics accounts.* Here we return to the brain's microarchitecture and dynamic profile in search of functional insights. Normal brains

display an information-processing ladder that leads ever forward from sensory neuronal populations through many intermediate populations and ultimately to populations of motor neurons. But there are also many axonal *back-projections* from populations higher and later on the ladder to populations lower and earlier. These 'descending' or 'recurrent' pathways introduce an intriguing variety of dynamic possibilities—such as self-modulation, selective attention, and autonomous activity—of interest to *all* of the functional accounts just scouted (Edelman 1993, Churchland 1995).

In particular, such a recurrent network can be configured so as to generate, autonomously, an unfolding trajectory in its neuronal activation space, a trajectory that partially *represents* an unfolding external reality. That trajectory can be continuously steered and edited by sensory input during the awake state; it can be left to wander freely during a disconnected 'dreaming' state, and it can be shut down entirely by the suppression of recurrent activity during 'deep sleep' (Llinás 2001).

These considerations hint that consciousness can come in a wide variety of degrees and flavours, depending on the character, the location, and the extent of such recurrent modulatory activity. Consciousness, some say, is like a light bulb—either it's on or it's off (Searle 1992). But perhaps not. A better analogy might be with light itself—which comes in endlessly different wavelength profiles and radiant intensities. Understanding light is a matter of grasping the relevant dimensions of variation. And just as genuine instances of light can vary widely along such dimensions, perhaps genuine consciousness may vary substantially from one species to another; or from one individual to another; or indeed, within a single individual over time.

*Extravagant accounts.* These deserve mention for reasons of history and completeness, but they look increasingly barren as sources of fecund research. *Dualism* posits an immaterial substance, distinct from the brain, in which consciousness inheres, forever beyond the explanatory reach of the physical sciences (Popper and Eccles 1978). *Epiphenomenalism* posits, not a substance, but a range of non-physical *properties* of the brain, similarly beyond any physical explanation (Jackson 1982, Chalmers 1996). And a recent suggestion posits *quantum-gravitational coherence* within the microtubules of the brain's axonal fibres as the hallmark or essence of consciousness, on grounds that blocking such coherence may explain how anaesthetics work, while the achievement of such quantum-level coherences may explain the existence of sound but non-algorithmic mathematical knowledge (Penrose 1994). Such options may be discussed in undergraduate classes, or in the media, but they play a negligible role in guiding empirical research.

Armstrong, D. (1981). *The Nature of Mind*.

- Baars, B. J. (1988). *A Cognitive Theory of Consciousness*.
- Bogen, J. E. (1995). 'On the neurophysiology of consciousness: an overview'. *Consciousness and Cognition*, 4.
- Chalmers, D. (1996). *The Conscious Mind*.
- Churchland, P. M. (1995). *The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain*.
- Crick, F., and Koch, C. (1990). 'Towards a neurobiological theory of consciousness'. *Seminars in the Neurosciences*, 2.
- Damasio, A. (1999). *The Feeling of What Happens*.
- Dennett, D. C. (1991). *Consciousness Explained*.
- Edelman, G. (1993). 'Neural Darwinism: selection and re-entrant signalling in higher brain function'. *Neuron*, 10.
- Jackson, F. (1982). 'Epiphenomenal qualia'. *Philosophical Quarterly*, 32.
- Leopold, D. A., and Logothetis, N. K. (1999). 'Multistable phenomena: changing views in perception'. *Trends in Cognitive Science*, 3.
- Llinás, R. (2001). *I of the Vortex: From Neurons to Self*.
- Lycan, W. (1987). *Consciousness*.
- Penrose, Roger (1994). *Shadows of the Mind*.
- Popper, K., and Eccles, J. (1978). *The Self and its Brain*.
- Searle, John (1992). *The Rediscovery of the Mind*.
- Singer, W. (2000). 'Phenomenal awareness and consciousness from a neurobiological perspective'. In Metzinger, T. (ed.), *Neural Correlates of Consciousness*.
- Taylor, J. G. (2001). 'The central representation: the where, what and how of consciousness'. In Colona, S. (ed.), *The Emergence of Mind*.