

Keith Sutherland

## *Why Do We Want to Open the Black Box?*

*Review Article*<sup>1</sup>

Several years ago the *Times Higher Education Supplement* published a debate on the causes of the recent explosion of interest in consciousness studies. According to Francis Crick consciousness was put back on the curriculum when Nobel Laureates (i.e. Gerald Edelman and himself) became active in the field, hitherto the province of New Agers and other assorted flakes. This led the neurologist Semir Zeki, in true Marxist fashion, to debunk Crick's great man theory of history and suggest a material alternative: according to Zeki the cause was the dramatic developments in imaging technology (PET, fMRI, etc.) The following week I pointed out that if all this gadgetry had been available to researchers in a more behaviouristically-oriented era they would have done the same experiments, but couched in terms of 'fractional antedating goal response' rather than the dreaded c-word. It was more likely that the re-emergence of consciousness studies occurred for sociological reasons: the students of the 1960s, who enjoyed a rich extra-curricular approach to 'consciousness studies' (even if some of them didn't inhale), are now running the science departments.

Susan Greenfield went up to Oxford in 1969 to read classics, promptly switched to psychology and is now Professor of Synaptic Pharmacology. Although *Brain Story* contains an interesting chapter on the effects of drugs, both prescribed and proscribed, on brain chemistry, she doesn't let slip whether her own interest in consciousness and the brain was ever influenced by recreational pursuits. However her background in the humanities is apparent throughout the book: Chapter Two starts with a Shakespearean meditation on the human condition, when she recalls the first time she dissected a human brain. But unlike the Prince of Denmark, who only had an empty skull to

Correspondence: Imprint Academic, PO Box 1, Thorverton EX5 5YX, UK

---

[1] Review of **Susan Greenfield**, *Brain Story* (London: BBC Worldwide, 2000, 208 pp. £17.99, ISBN 0-563-5510-8. An abridged version of this review was first published in the *Times Higher Education Supplement*.

contend with, Greenfield was intrigued whether a tiny sliver of cortex scraped under her fingernail (if she had not been wearing gloves) might contain a particular memory, fear or dream.

The interweaving of such anecdotes with popular scientific material shows that this book is intended for a wide audience. *Brain Story* is delightfully written, lavishly illustrated, carefully edited and is clearly targeted at anyone with even a casual interest in the workings of the brain. Comparatively well-known terms like long-term potentiation (LTP) are apologetically introduced as ‘jaw-breaking’. But why should Joe Public be in the slightest bit interested in the difference between LTP and LGN? Given that most primary school children know that our experiences are somehow or other caused by brain events, why should we want to bother opening up the black box? Although it is obviously important for a practising neurosurgeon to know which cortical areas are involved in, say, speech production, does this information represent any progress beyond nineteenth-century phrenology as far as the general public are concerned?

In fact there are several good reasons why this book should be read by a wide range of people (not just those who, like my mother, switched off the accompanying television series because they were so irritated by the intrusive incidental music). Over the last few decades the public has been presented with a highly misleading caricature of brain science by the media and *Brain Story* goes a long way towards correcting these distortions.

Ask any non-specialist (not to mention a hard rump of old fashioned cognitive scientists) about the mind and chances are you will hear some version of the old yarn ‘mind is to brain as computer software is to hardware’. Although this myth has its origins in the science-envy that led psychology to make its Faustian pact with the artificial intelligence industry, neuroscientists have not been entirely free of culpability. In the 1960s no less a person than the Nobel Prize-winning physiologist John Eccles fell victim to the highly infectious meme that the neuron was some kind of binary switch that only permitted two ‘digital’ states (excited and inhibited). Greenfield pours scorn on this by outlining more recent evidence on the modulatory role of neurotransmitters, along with a discussion of Rodolfo Llinás’s extraordinary discovery that dendrites from the cerebellum, usually viewed as humble one-way conduits, are themselves involved in a kind of learning process. Even without considering some of the more exotic theories about microtubules and the cytoskeleton, neurons are exquisitely complex and it is a serious error to represent them as simple digital switches. Also brains are integrated into bodies and are subject to all manner of hormonal and visceral influences. ‘To model the brain in its entirety, one needs to model a body too’ (p. 105).

Bodies are also situated in a social and cultural milieu. Although pharmacologists normally restrict themselves to studying how, say, amphetamine increases arousal by boosting the effect of dopamine in the brain, Greenfield reports research that shows that the same drug, the same arousal, can produce markedly different emotions, depending on the social context. Experimental subjects who were expecting to be given amphetamine enjoyed its effects but subjects who were (untruthfully) told that they had been given a placebo simply felt anxious. It is

refreshing to find a pharmacologist who is as strongly opposed to the temptation to explain sophisticated states of mind ‘exclusively in terms of mere molecules’.

One of the reasons for the dominance of the computational theory of mind is the cross-infection between work in the neurophysiology of vision and computer vision systems. This has led to the simplistic view that the human visual system operates on a one-way input-driven model. Greenfield points out that vision is a highly active system, replete with top-down feedback loops and even goes so far as to endorse Llinás’s radical view that there is no essential difference between waking and sleeping. In both cases the brain actively constructs the world — waking is just dreaming with the eyes open and without sensory constraints.

Seeing as no-one has the faintest idea how human memory works, the temptation has been to borrow words like ‘store’, ‘retrieve’ and ‘memory trace’ from the computer science literature. But there is no proven parallel between human and computer memory. Computer memory is localised whereas, as Wilder Penfield famously discovered, there is no reliable correlation between the cortical area stimulated and the specific memory evoked. Even the surgical removal of a complete hemisphere rarely ablates specific memories — if cognitive science is in need of an appropriate metaphor for memory then holography would be a more fruitful source. Hippocampal damage tends to affect the time threshold for memory formation and recall rather than specific memories. A study by some NYU researchers published recently in *Nature* (Nader *et al.*, 2000) showed that, unlike computer memory, long-term memory in human subjects was surprisingly labile (chemically unstable) — every time a memory was recalled it was modified and ‘reconsolidated’.

If the authors of the recent Department of Trade and Industry ‘Foresight’ report had taken the trouble to read Greenfield’s book they would not have come out with the widely-trailed statement that ‘neuro-chemical technology may provide greater access to the memories of the living and possibly deceased’. This has led to the *Cold Lazarus*, science fiction prediction that the pathologist of the future will be able to access the ‘memory traces’ of the murder victim, leading the Canadian ‘futurologist’ Frank Ogden to claim that

[Memory] is just information like anything else. The mind is similar to a computer. And even when you erase something on a computer, by putting it in the trash, it’s still there.

The main value of Greenfield’s book is as an antidote to futuristic trash like this. She is disinclined to accept the operational definition of machine consciousness that is examined in the Turing Test,<sup>2</sup> preferring arguments about intuition and common-sense derived from Roger Penrose. She holds out some hope that artificial neural networks may provide some useful insights into how brains operate, but I suspect this is partly because the biologically-plausible terminology of

---

[2] Turing’s original paper was concerned with machine *intelligence*, rather than consciousness, but given that it was written at the height of the behaviourist era, it is hard to imagine what other term he might have used.

ANNs (genetic algorithms and the like) and their emphasis on ‘learning’ conceals the binary beast at their heart.

The computational model of the mind–brain has recently been combined with a Darwinism-inspired theory of origins to form the new hybrid discipline of ‘evolutionary psychology’. According to evolutionary psychologists like Steven Pinker there is no such thing as ‘general intelligence’, rather what we like to call ‘the mind’ is a bolted-together system of modules that developed in the ‘environment of evolutionary adaptedness’ (EEA) in order to solve specific survival-related problems. Evolutionary psychology has been much encouraged by the evidence on the modular nature of cognitive functions uncovered by neuropsychologists in lesion studies, much of which is reported in this book.

Larry Weiskrantz’s famous ‘blindsight’ patient GY is ‘outed’ in this book as Graham Young. We can leave aside the issue of whether the reports of a single patient (who has become something of an expert on the topic and a ‘superstar in the field of consciousness research’) can really be trusted. Fortunately, most of the other examples of selective brain damage reported in the book rely far less on fine distinctions like whether a subject is dimly aware or only guessing. The evidence from patients like Gisela Leibold, who suffers motion blindness as a result of very specific stroke damage, and Lincoln Holmes, who cannot recognise faces, would tend, on superficial examination, to support the ‘building-block’ hypothesis of evolutionary psychology.

Although this all makes for good prime-time television, in the book Greenfield is a lot more cautious about extrapolations from lesion studies — we should be ‘wary of inferring function from dysfunction’ (p. 168). If damage to Broca’s area causes language impairment, this does not support the claim that Broca’s area is the ‘language centre’:

The fallacy of this conclusion is obvious if you imagine the same logic applied to a radio — if you took a valve out of a radio and it started to howl, that would not mean that the function of the valve was to inhibit howling (p. 168).

There is a growing wave of scepticism about the research methodology in cognitive neuroscience — the correlative study of brain images and behavioural tests — with its built-in theoretical bias towards localisation. Perhaps this is nothing more than an ‘easy way out for experimental design’ (Uttal, 2000). One survey (Grafman *et al.*, 1995) devotes seven pages to the various cognitive processes that have been associated with the frontal cortex. Pulvermüller (1999) has pointed out that the cognitive processing of word meanings has been ‘located’ in all of the major lobes of the brain. So are the ‘components of mental activity’, the mental modules, anything more than *a priori* or *ad hoc* hypothetical constructs, cobbled together for experimental purposes? Does the elaborate processing required to extract results from the raw scanning data not cast serious doubts on the validity of the findings? (Uttal, 2001).

No doubt Greenfield would agree with Jerry Fodor, who concludes in his new book *The Mind Doesn’t Work That Way* (a direct response to Steven Pinker’s *How the Mind Works*), that the evidence for modularity just isn’t available. The trouble

is, Fodor concludes, if the modular theory is wrong we simply do not possess a better theory, so it's back to the drawing board.<sup>3</sup>

Greenfield proffers her own dynamic theory as a candidate, according to which transient assemblies of tens of millions of brain cells — distributed across many different brain areas — compete with each other to generate moments of consciousness.<sup>4</sup> According to Greenfield, consciousness is not a 'magic switch', rather a dimmer switch, a continuum. There is no sudden transition (in phylogeny or ontogeny) to sentience, only a question of degree.

She uses evidence from the incremental ablation of consciousness under anaesthetics to support her theory, speculating that deeper levels of anaesthesia are associated with smaller neuronal assemblies. To Greenfield, 'increased' consciousness means larger, more complex neuronal assemblies, in which sensory input, emotion, rationality, and the richness of past experience all combine. This leads her to conclude that dreaming and 'being trapped in the present moment' are lower forms. However, some of our most vivid experiences occur during dreams and countless systems of esoteric philosophy value the 'experience of the present moment' or the 'pure' (contentless) conscious event as the most highly developed form of conscious experience. But perhaps there is no conflict here as such philosophies often equate the onset of (human, linguistic) consciousness with the Fall and seek to offer the path back to the Garden of Eden.<sup>5</sup>

Unfortunately for Greenfield there is no evidence to support her theory of neuronal groupings as, just as Victorian photography could not register moving objects, current imaging technology is too slow to capture such transient neural ensembles. Moreover, although the theory might well explain the differing qualities of conscious events and how a particular assembly may achieve victory in the competition for consciousness, it does not address the so-called 'hard problem' as to why a brain event of any shape and form should give rise to conscious experience (see below).

Greenfield cites Benjamin Libet's discovery — a pinprick to the hand only takes 20 milliseconds to reach the brain, but a further half second to generate conscious experience — as evidence for the time needed to recruit a large enough neuronal assembly. As human subjects can respond to stimuli much faster than this, this leads her to question, with psychologist Jeffrey Gray, what the purpose of consciousness is. The example usually given is the Centre Court at Wimbledon where, we are led to believe, unconscious zombies would be just as effective as Pete Sampras or Venus Williams. The reason given is that the ball is returned before it is consciously perceived. However, it would be just as plausible to say that through watching (consciously) the angle of the serve and anticipated

[3] On a similar vein, Jonathan Horton, writing the millennium essay in *Nature*, concludes that 'the brain still resists researchers' attempts to divide it into neat parcels' (Horton, 2000).

[4] It is important to distinguish this from the obsolescent idea of 'mass action' in the nervous system.

[5] This has been nicely portrayed as the 'Krishnamurti Dog Problem'. 'Krishnamurti commented once that for the duration of an hour's walk not a single thought entered his mind — this raises for us the problem of his dog, for whom we can assume the same state for the period of their walk. What then is the difference in consciousness between these two beings?' (King, 1996).

trajectory the skilled player is able to anticipate the timing of the return. There is no evidence to prove Greenfield's claim that 'returning a tennis ball is a complex process — all done entirely subconsciously' (p. 182).

### **Ethics and Religion**

The second reason why this book is so valuable is because much of the debate on scientific ethics of late has been devoted to issues of consciousness. Greenfield, who as Director of the Royal Institution is Britain's best-known neuroscientist, has made a number of statements in the public domain, expressing doubt over the need for brain-dead organ donors to receive anaesthetics prior to organ removal. On the other hand she has added her voice to the growing concerns over whether foetuses undergo pain during abortion. The spectrum of opinion on the threshold of consciousness now ranges from panexperientialism at one extreme — where all matter is 'proto-conscious' — to the view that only creatures that possess language are sentient. Technological developments will mean that issues concerning the onset of consciousness will become an increasing focus of medical ethics and Greenfield's book provides a useful introduction to ethical issues that should be of interest to all citizens.

However, I think the main reason that lay people might want to read this book is that, given the decline in the status and plausibility of mainstream religion, the public increasingly looks to science to fill the gap. Can this book fulfil such expectations?

The first episode of the television series was devoted to the scientific examination of religious experience and was filled with statements like 'scientists believe that all experience can be ultimately explained by the functioning of nerve cells in the brain' (Francis Crick's astonishing hypothesis that 'you're nothing but a pack of neurons'). Greenfield appeared to take a great delight in explaining both Van Gogh's creative use of colours and religious and mystical experience as some form of temporal lobe epilepsy. But does this tell us anything other than the pre-theoretical intuitions of a philosophical materialist (Democritus is one of the few philosophers who make it into this book)? Are attempts to investigate the truth claims of religion through the study of the brain any better founded than William Paley's attempts to prove the existence of God through the argument from design?

If alien scientists from the planet Zarg were asked to reverse engineer a television set, no doubt they would conclude that the images on the screen were generated by the transistors and other internal components and that the aerial was some sort of cooling device. Even if Ockham's razor is invoked against such arguments, human experience involves a sentient creature acting in a cultural and social context, rather than a brain in a vat. No amount of knowledge of the detailed workings of the brain will ever be able to 'explain' human experience.

Greenfield acknowledges the so-called 'hard problem' of consciousness — 'how a subjective, inner world is generated from a lump of neuronal sludge' (p. 173) — as formulated by philosophers like David Chalmers. But she then

seeks to get round the problem using John Searle's idiosyncratic notion (I can think of no other philosopher who would agree with it) that the problem of phenomenal consciousness and the problem of free will are one and the same. She then executes a quick segue to a discussion with Benjamin Libet on his findings that the notion of voluntary action (the free will problem) is an illusion as experimental evidence shows that the brain develops a 'readiness potential' a full half second prior to the experience of a volition. Thus our so-called free will is just an epiphenomenon of deterministic brain processes. 'If consciousness and free will are mere illusions, where does this leave personal responsibility and accountability?' (p. 184).

Indeed. However Greenfield has left out half of the story. In the *JCS* special issue that he co-edited with Anthony Freeman and myself (Libet *et al.*, 1999), Libet claims that we can always veto any 'voluntary' act and there is no evidence to suggest that the veto is anything other than an act of free will.<sup>6</sup>

I'm afraid misrepresenting Libet and Searle will not get Greenfield off the hook. The seeming intractability of the 'hard problem' of consciousness has driven many hard-headed scientists and sober scholars to metaphysical extremes. If (as Greenfield claims) there is no sudden onset of consciousness, no magic switch to flip, and if (as she agrees) we have no idea at all how a lump of meat might generate experience, then panexperientialism — the idea that proto-consciousness goes 'all the way down' — is a way out of this impasse that is being reluctantly embraced by a growing number of scholars. Greenfield is happy to cite Roger Penrose's work on non-computability in support of her argument against artificial intelligence, but she fails to cite the controversial Penrose-Hameroff theory that experience might be a fundamental property of space-time.

No doubt these metaphysical problems remain firmly locked away in the closet in order not to frighten the horses. If, as Greenfield declared in her interview with the religious affairs correspondent of a Sunday newspaper, 'the brain is her religion', then the last thing the new convert needs is a doctrinal dispute among the High Priests.

## References

- Fodor, J. (2000), *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology* (Cambridge, MA: The MIT Press).
- Grafman, J., Partiot, A. & Hollnagel, C. (1995), 'Fables in the prefrontal cortex', *Behavioral and Brain Sciences*, **18**, pp. 349–58.
- Horton, J.C. (2000), 'Boundary disputes', *Nature*, **406**, August 10, p. 565.
- Libet, B., Freeman, A. and Sutherland, K. (1999), *The Volitional Brain: Towards a Neuroscience of Free Will* (Exeter: Imprint Academic).
- King, M. (1996), 'From Schroedinger's cat to Krishnamurti's Dog: mysticism as the first-person science of consciousness', *Consciousness Research Abstracts: Toward a Science of Consciousness 1996*, pp. 141–2 (Exeter: Imprint Academic).
- Nader, K., Schafe, G.E. and LeDoux, J.E. (2000), 'Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval', *Nature*, **406**, August 17.

[6] Libet informed me that his interview for Greenfield's book was a long one and this aspect was conveniently edited out.

- Pinker, S. (1999), *How the Mind Works* (New York: W.W. Norton & Co.).
- Pulvermüller, F. (1999), 'Words in the brain's language', *Behavioral and Brain Sciences*, **22**, pp. 253–336.
- Uttal, W.R. (2000), 'On the limits of localization of cognitive processes in the brain', *MIT CogNet*: [http://cognet.mit.edu/bboard/ed-com-msg.tcl?msg\\_id=00004c&latest=1](http://cognet.mit.edu/bboard/ed-com-msg.tcl?msg_id=00004c&latest=1)
- Uttal, W.R. (2001), *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain* (Cambridge, MA: MIT Press, forthcoming).