EMBO reports

viewpoint

Human agency in the neurocentric age

Philosophers and scientists resort to dualistic explanations to reconcile the age-old dichotomy between determinism and 'free will', but agency is an integral part of human biology

Steven P.R. Rose

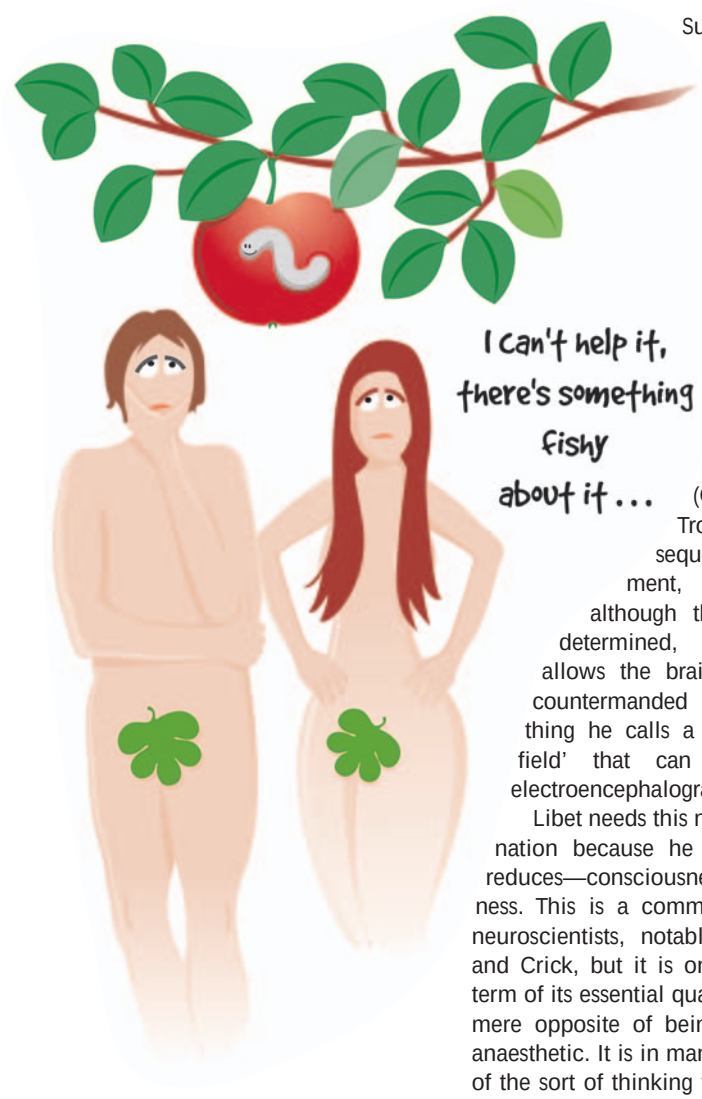
"Freely they stood who stood, and fell who fell.
Not free, what proof could they have giv'n sincere
Of true allegiance, constant faith, or love…
I formed them free, and free they must remain,
Till they enthrall themselves: I else must change
Their nature, and revoke the high decree
Unchangeable, eternal, which ordained
Their freedom…" (Milton, 1667)

The debate on whether humans have free will or a predetermined fate has been a central theme for all religions of the Book, but no more strongly than within Christianity and its successive, secular philosophies. In his epic poem *Paradise Lost*, John Milton's insistence on human freedom to act—whether good or evil—stands in sharp contrast with predestinarian sects who argued that, even at birth, humans were already pre-assigned to hell or heaven. Within the Christian tradition, this debate is resolved by various forms of dualism, most clearly expressed by the French philosopher René Descartes (1596–1650). For Descartes, animals were mere mechanisms, whose cogs and pulleys were built of meat. It was only with humans that God had inserted a soul, which, interacting with the meat-machine through the pineal gland, allows conscious thought and willed action, creating saints or sinners. This Miltonian understanding of freedom was later echoed by the secular existentialist philosopher Jean-Paul Sartre (1905–1980): "...there is no human nature...Man simply is.…he is what he wills...one will never be able to explain one's action by reference to a given and specific human nature; in other words, there is no determinism—man is free, man is freedom" (Sartre, 1946).

This notion that, by virtue of being human, humans possess agency—the possibility of acting based on intention—has been a central tenet of Western thought and philosophy. As agents, humans can act freely; my actions are governed solely by my choice. I am 'free' to lift my arm above my head unless I am constrained by some external or internal force: by a prison guard threatening me, or by some muscle-wasting disease. I am 'free' to dine at the Ritz or sleep under Waterloo bridge—depending on the amount of money in my bank account. However, this notion of human freedom has been systematically under attack by the biosciences, no more sharply than in Dawkins' claim that "we are survival machines—robot vehicles blindly programmed to preserve the selfish molecules known as genes" (Dawkins, 1976). Such genetic determinism has recently been joined by a neuronal determinism—what Racine *et al* (2005) refer to as neuro-essentialism, and what I refer to as neurogenetic determinism (Rose, 1995). Crick's aphorism "You're nothing but a pack of neurons" (Crick, 1994) neatly encapsulates this argument. However, in the closing sentence of *The Selfish Gene*, Dawkins argues that "only we have the power to rebel against the tyranny of our selfish replicators", whereas Crick's *Astonishing Hypothesis* concludes with a discourse on Free Will (his capitals), which he locates, rather mischievously, to a portion of the brain known as the anterior cingulate sulcus.

But Dawkins' and Crick's arguments fall victim to the very same paradox that has plagued Western philosophy for centuries: if we are nothing but robot vehicles blindly programmed to preserve our selfish genes, from whence comes the 'we' with the power to rebel, and what allows the cingulate to contain 'free will'? If it is programmed in our genome and the epigenetic cascade through which our cingulate becomes wired up (Robert, 2004), how can it be free? The emphasis on epigenesis here is important, for there is no conceptual difference between claiming that we are determined through our genes or that we are determined through our childhood experiences and the socio-economic context in which we are reared. In each case, free will would seem to be nothing other than a "user illusion" (Nørretranders, 1998)—an epiphenomenon to be dismissed summarily, as Churchland does, as "folk psychology" (Churchland, 1995). For those troubled by this paradox, it would be more honest to take the Cartesian escape route favoured many years ago by the neurophysiologist Sir John Eccles, a committed Catholic who argued that the soul could influence action by tinkering with synaptic transmission in what he called "the liaison brain" in the left hemisphere (Eccles, 1964), or that offered by Roger Sperry with his plea for "downward causation" (Sperry, 1985). Indeed, this is in essence what Dawkins does. By separating the 'we' who can rebel against our genes and our brain, this avowed materialist becomes a Cartesian dualist in the laudable interest of preserving human agency.

I find this abdication unsatisfactory, and instead want to insist that our sense of freedom to act, of possessing agency, emerges inevitably from our biological nature. Like all other living organisms, humans are the result of an autopoietic process (Maturana & Varela, 1980)—our self-creation from the raw materials of our genes and environment (Rose, 2005a). We are, however, qualitatively different because our self-creation is embedded in a social, historical, cultural and technological environment, and as a



consequence of our social nature, we are also moral beings. It is from this context that 'free will' emerges, not from a soul or a particular region in the brain.

The discussion of free will is inextricably linked with that of consciousness. If it is our brain that assesses input and determines output, neuroscientists ask, what is the purpose of consciousness, as if the property of human consciousness could be separated from other brain processes in some version of neo-Cartesianism (Searle, 2004). The experiments most often cited in this regard are those of Libet (2004), who showed that during 'voluntary' hand movements, a readiness potential could be measured some 300 milliseconds (ms) before the subjects claimed to be consciously 'aware' of their decision to move the hand. The brain had already 'made the decision' and the person's action merely followed. Such experiments reflect the observation that skilled tennis players or batsmen adjust to incoming balls and play their strokes too rapidly for 'conscious' decision-making to occur. Some have even argued that such apparent spontaneity even makes for better decisions (Gladwell, 2005). Troubled by this consequence of his experiment, Libet argues that although the action may be determined, the 300-ms gap allows the brain's decision to be countermanded by virtue of something he calls a 'conscious mental-field' that can be detected by electroencephalography.

Libet needs this neo-Cartesian explanation because he equates—or rather reduces—consciousness to mere awareness. This is a common strategy among neuroscientists, notably by Koch (2004) and Crick, but it is one that empties the term of its essential qualities, making it the mere opposite of being asleep or under anaesthetic. It is in many ways reminiscent of the sort of thinking that Dennett (1991) memorably characterized as 'the Cartesian theatre', in which possible choices for action are presented to some homunculus in the brain, which then instantaneously decides and instructs the relevant brain processes to proceed. The processes by which 'we' assess the potential outcomes of our actions in Libet's 300-ms gap may simply not be available to introspection unless we specifically train ourselves to do so, as when Yoga practitioners insist that we become aware of bodily processes that normally occur without awareness, such as breathing or heartbeat.

> … if we are nothing but robot vehicles blindly programmed to preserve our selfish genes, from whence comes the 'we' with the power to rebel…

Libet's explanation also trivializes the complexities of the processes engaged in conscious decision-making. My lab has been using magneto-encephalography to map the cortical processes involved in everyday decision-making, after asking subjects to choose between alternative consumer products in a virtual visit to a supermarket (Bräutigam *et al*, 2001). It typically takes more than two seconds for a subject presented with a choice between, say, three brands of breakfast food, to indicate their preference by pressing a key. During the first second of the decision-making process, a wave of activity (first detectable at around 80 ms) in the visual cortex passes through the inferotemporal cortex and Broca's area (at around 800–900 ms) to the right parietal, which lights up only if the subject has a strong preference for one of the items.

In this experiment the choice is entirely free in the sense that it is unconstrained—subjects do not have to concern themselves with the relative costs of the items, for instance. However, the subject's preference will be based on past experience, itself presumably the result of earlier 'free' choices. Asked when they became aware that they had made their choice, subjects would presumably, on the Libetian model, indicate some time around 300 ms after the parietal became active, but still well in advance of the key press. However, my point is that conscious decision-making involves the entire sequence up to the key press. There is no paradox about it being both 'free' and yet determined (Lipton, 2004), nor is this freedom impaired by the fact that we can describe the brain processes (Midgley, 2004). This is also true of the much larger decisions we make throughout our lives: to accept a new job offer, to move house, or to get married etc.

So let us explore the limits of neuroscience's ability to reduce mentation—'folk psychology'—to mere brain processes. I take for granted that my thoughts, intentions and motives—those key features of agency—are emergent properties that arise from specific states of my brain and body. Consider the following thought experiment. Imagine that I have all the technologies and information-processing power that neuroscientists can dream of, and build a machine—let's call it a cerebroscope, a term I believe was invented by the information scientist and explicitly Christian

anti-determinist Donald Mackay—that can report the activities at any one time of all 100 billion neurons in my brain (Rose, 2005b). Now consider the experience of 'seeing a red bus coming towards me'.

> …our sense of freedom to act, of possessing agency, emerges inevitably from our biological nature

The cerebroscope will record and integrate the activity of many neurons in the visual cortex, those that are wavelength sensitive and report red, those that report directional movement, edge-detecting neurons, neurons tuned to binocularity and so forth, all of which combine to create a mental image of an object moving towards me with a speed I can estimate from the rate of change that the image subtends on my retinae. Acoustic information is also included, so I can register the engine noise of the approaching bus. But how do I know that the noise is of an engine, or that the object is a bus? Some other neural activity scans my recognition memory, which defines the object as a bus, and the noise as that of an engine. Is seeing this bus a good or a bad thing? If I am waiting for it, a good thing; if I am crossing the road, a dangerous thing. Then 'I' must 'decide' how to act: prepare to enter or jump out of the way? The appropriate muscles must be engaged, blood circulation adjusted, and so on and so forth.

The cerebroscope will enable an observer to record all this neuronal activity and entitle him to say that the total sum of this activity represents my mental processes of encountering the bus. So, there is an identity between mind processes and brain/body states and dynamics. Now suppose the cerebroscope stores all this information. Then at some later time, an experimenter asks the machine to present the data and to deduce from the neural activity the thought and action processes that it represents. Could it interpret all the data and print out a statement 'this brain is experiencing a red bus coming towards the body in which it is embedded and that it is in danger of being run over by it'? Or, to take another example beloved of the new enthusiasts for consciousness studies, would feeding the data into the artificial brain of a zombie result in the zombie catching the bus (Malik, 2000)?

The answer seems to me to be 'no'. The firing pattern of any particular neuron is very much dependent on its history. Plasticity during development may mean that even the wavelength to which any particular neuron is sensitive may vary from individual to individual, so what ends up as one person's 'red' neuron may be another person's 'blue' one. Even more certain is that whatever is the pattern of neural firing and connectivity in my inferotemporal cortex that corresponds to my recall or recognition memory of a bus, it is not the same as the pattern in yours, even though the outcome—recognizing a bus—will be the same in both cases. This is because your and my experience of buses, and how we store that experience, is inevitably unique to each of us.

It follows that for the cerebroscope to interpret a particular pattern of neural activity as representing my experience of seeing the red bus, it needs access to my entire neural and hormonal life history. Then and only then might it be possible to translate the neural information back into a statement about my mentation and experience. However, this would still only be true if there were a one-to-one relationship between the history and present state of my neurons and my mental activity. And this we simply do not know. There may be several histories of neural firing patterns from my conception to the present time that could be interpreted as experiencing a red bus coming towards me—and equally there might be an indefinitely large number of experiences that could be inferred from any particular pattern. That is, whereas the cerebroscope may record a specific pattern of brain states associated with my bus-catching experience, the relationship may simply not be transitive.

The problem becomes more acute when one considers not just the interpretation of current sense data based on past experience, but more abstract problems, of the sort raised 16 centuries ago by St Augustine of Hippo (354–430). How, he asked, could the brain contain "abstract propositions, the principles of numbers and dimensions…false arguments…the idea of god?" (Pine-Coffin, 1961). As Dehaene (1997) has shown, there are regions of the brain that become active when people are asked to solve mathematical problems, and there are distinct differences in brain activity between those with weak and those with strong mathematical skills. Others have spoken of identifying a 'god centre' in the brain—presumably a region that is active when a believer thinks about his or her personal deity. Perhaps this shares a location with Augustine's 'false arguments', as there are also claims of the ability to distinguish between 'true' and 'false' memories on the basis of brain signals (Schacter, 1999).

So let us bring back the super-cerebroscope and focus it on the brain of someone trying to decide whether an argument is true or false. Once again, we expect all sorts of brain regions to light up as some proposition is examined, compared with related propositions extracted from memory, and so forth. The cerebroscope will in due course also register the final decision as to the truth or falsity of the proposition. But would it be able to detect the actual content of the argument leading to the conclusion? I suggest not; the cerebroscope is at the limits of its powers in identifying the brain regions that enable the mental process involved in the argument. It is at this point, I suggest, that neuroscience may be reaching its theoretical limits in understanding the brain in order to explain the mind.

The sociologist Hilary Rose (2000) has posed a further limit to cerebroscopy. In Europe and the USA, a profound change has occurred in the relationships of men and women over the past 30 years, as a consequence of the rise of feminist consciousness. The way that many women—and therefore, of necessity, many men—perceive their relationships has been altered. We must assume that this transformation has caused significant changes in both neural and hormonal processes in the body, detectable by our cerebroscope. But would the researcher reading the cerebroscope records be able to interpret these changes as indicative of the rise in feminist consciousness? Even to pose the question seems to indicate the limit of the cerebroscope's potential power. To interpret its readings, the cerebroscope would need to have information about the social and cultural context in which the individual and her brain was embedded. And this is precisely the point: our mental activity, our consciousness, cannot be reduced in the formal, philosophical sense of reduction (Rosenberg, 1985) to just what is happening in our brain at any moment. Mentation and the actions that follow always depend on the larger context (Donald, 2001).

> Mentation and the actions that follow always depend on the larger context

Some years ago at a symposium on reductionism in the biological sciences, I clashed with the philosopher Thomas Nagel who suggested that although higher-level accounts, such as mentalistic ones, could describe a phenomenon, only the 'lower-level' reductionist ones could explain it. I disagree. In many cases, lower level accounts are descriptive whereas the higher-level ones are explanatory. Consider, for instance, memory. Research over many decades has produced an account of the molecular cascade occurring during memory formation. However, this summary does not explain the memory, it merely describes the brain events involved in making it. For an explanation, look to psychology or ethology, not neuroscience (Rose, 2004). However comprehensive the cerebroscope's record of the neural activity taking place when I experience the sensation of being angry or in love, drafting this sentence or designing an experiment, the account will only be descriptive, not explanatory.

Human agency is a statement about the autonomy to act as a person, not as a robot, and the actions of that person are not reducible to the properties of a pack of neurons. This is not just a debate between the religious and the secular, or between neuroscientists and philosophers. It is because we have agency that we are regarded to be responsible for our actions, in both the moral and the legal sense. It is therefore not surprising that it is in the context of the law where recent neuroscientific advances and claims are being taken most seriously (Zeki & Goodenough, 2004; Garland, 2004; Botkin *et al*, 1999). The legal system has long puzzled over whether a person found guilty of a crime was acting with *mens rea*—sound mind—or was suffering from some constraint, internally or externally generated, which diminished responsibility. Thus, in English law, children cannot be held responsible for their actions below the age of ten, as they are presumed not to be mature enough to distinguish between right and wrong.

Even in adulthood, someone who has killed another should not be convicted of murder "if he was suffering from any abnormality of mind (whether arising from a condition of arrested or retarded development of mind or any inherent causes or induced by disease or injury) as substantially impaired his mental responsibility for his acts or omissions…" (Sedley, 2004). Under such circumstances it would not be 'me' that freely performed the act; I was under instruction from my brain (neuro-essentialism), which in turn may have been impaired by any juridically accepted reason implied in the quotation above. The law thus makes a Cartesian distinction between 'the man' and 'the disease'. On the basis of claims that the presence of an abnormal allele of monoamine oxidase A is associated with aggressive behaviour (Brunner *et al*, 1993; Caspi *et al*, 2002), a genetic defence, or plea in mitigation, for a convicted murderer has been attempted, although I believe, not accepted by a US Court. An analogous case revolved around whether a man who shot several co-workers and then committed suicide was not responsible by virtue of the fact that he was taking the prescribed drug Prozac®, known to be associated sometimes with violent or suicidal behaviour (Cornwell, 1996; Healy, 2004). But a plea of acting criminally under the influence of alcohol is not acceptable, as the courts regard getting drunk as a voluntary act for which one is responsible.

> It is because we have agency that we are regarded to be responsible for our actions, in both the moral and the legal sense

However, if agency can be impaired by immaturity or injury to the brain, it must follow that 'normal' agency too is brain- and gene-dependent. It has been argued that a predisposition to drink, or to act impulsively and hence criminally, is heritable. Even if it were not, to all our acts, intentional or not, there must be a corresponding brain state. Before that act, there must also correspond some genetic or brain trait that determines it—some feature of neuroanatomy or quirk of neurotransmitter or neuromodulator levels, as in untreated disturbances of dopamine metabolism in childhood, which are said to predict future criminality (Barkley, 2002).

As neuroscientists, we may regard the efforts of the courts to distinguish between situations in which a person is 'free to act' and therefore culpable from those in which they are constrained and therefore not responsible for their actions, as scientifically nonsensical, creating untenable distinctions that scientific advance and criminal defence lawyers will constantly seek to erode. But, I suggest that legal common sense, which attributes responsibility in all cases except where an action is clearly over-determined by factors outside the agent's control, also makes better philosophical sense, and may help us avoid some of the more extreme reductionist propositions to which our science has become addicted. If this makes common sense so far as the law is concerned, perhaps it should help us in other, less practical, areas as well.

> …I suggest that legal common sense… also makes better philosophical sense, and may help us avoid some of the more extreme reductionist propositions to which our science has become addicted

In trying both to redefine and to rescue the concept of human agency, I reject the implicit dualism that separates 'me' from my brain—or more accurately, from my brain/body system—while also refusing to consign agency to the dustbin of false ideas. The conceptual confusion that surrounds determinism and free will is deeply embedded in our way of thinking because neuroscience, like all modern science, has developed from within the Judaeo-Christian traditions. In truth, we live at the interface of multiple determinisms. My freedom to dine at the Ritz is enabled by my credit limit, and to lift my arm above my head by whether I am suffering from myopathy or a mere frozen shoulder. For every action we take, it is possible to define causes at many levels, from antecedent neural events to cultural norms and the financial constraints of a market economy. The important scientific question then is to know at which level it is appropriate to seek an over-determining cause. To understand and hopefully to treat Alzheimer's disease, we need to know about the biochemistry of the amyloid precursor protein, but it would be folly to try to explain the causes of the invasion of Iraq in 2002 in terms of fluctuations in transmitter levels in US President Bush's brain. We are, to summarize my argument, free to act and to shape our own future, although not in circumstances of our own choosing.

"The mind is its own place, and in itself
Can make a Heav'n of Hell, a Hell of Heav'n."
(Milton, 1667)

ACKNOWLEDGEMENTS

I thank Pat Bateson, Peter Lipton, Mary Midgley and Stephen Sedley for helpful comments on an earlier draft of this paper.

REFERENCES

Barkley RA (2002) International Consensus Statement on ADHD. *Clin Child Fam Psychol Rev* **5:** 89–111

Botkin JR, McMahon WM, Francis LP (eds) (1999) *Genetics and Criminality: The Potential Misuse of Scientific Information in Court*. Washington, DC, USA: American Psychological Association

Bräutigam S, Stins JF, Rose SP, Swithenby SJ, Ambler T (2001) Magnetoencephalographic signals identify stages in real-life decision processes. *Neural Plast* **8:** 241–254

Brunner HG, Nelen M, Breakefield XO, Ropers HH, van Oost BA (1993) Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. *Science* **262:** 578–580

Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig IW, Taylor A, Poulton R (2002) Role of genotype in the cycle of violence in maltreated children. *Science* **297:** 851–854

Churchland PM (1995) *The Engine of Reason, the Seat of the Soul: a Philosophical Journey into the Brain*. Cambridge, MA, USA: MIT Press

Cornwell J (1996) *The Power to Harm: Mind, Medicine, and Murder on Trial*. London, UK: Penguin

Crick F (1994) *The Astonishing Hypothesis: the Scientific Search for the Soul*. New York, NY, USA: Scribner

Dawkins R (1976) *The Selfish Gene*. Oxford, UK: Oxford University Press

Dehaene S (1997) *The Number Sense: How the Mind Creates Mathematics*. Oxford, UK: Oxford University Press

Dennett DC (1991) *Consciousness Explained*. Boston, MA, USA: Little, Brown and Co.

Donald M (2001) *A Mind So Rare: the Evolution of Human Consciousness*. New York, NY, USA: Norton

Eccles JC (1964) *The Physiology of Synapses*. Berlin, Germany: Springer

Garland B (ed) (2004) *Neuroscience and the Law: Brain, Mind, and the Scales of Justice*. New York, NY, USA: Dana Press

Gladwell M (2005) *Blink: the Power of Thinking Without Thinking*. New York, NY, USA: Little, Brown and Co.

Healy D (2004) *Let Them Eat Prozac: the Unhealthy Relationship between the Pharmaceutical Industry and Depression*. New York, NY, USA: New York University Press

Koch C (2004) *The Quest for Consciousness: a Neurobiological Approach*. Denver, CO, USA: Roberts & Co.

Libet B (2004) *Mind Time: the Temporal Factor in Consciousness*. Cambridge, MA, USA: Harvard University Press

Lipton P (2004) Genetic and generic determinism: a new threat to free will? In Rees D, Rose S (eds), *The New Brain Sciences: Perils and Prospects*. Cambridge, UK: Cambridge University Press

Malik K (2000) *Man, Beast and Zombie: what Science Can and Cannot Tell us about Human Nature*. London, UK: Weidenfeld & Nicolson

Maturana HR, Varela FJ (1980) *Autopoiesis and Cognition: the Realization of the Living*. Boston, MA, USA: D. Reidel Publ Co.

Midgley M (2004) Do we ever really act? In Rees D, Rose S (eds), *The New Brain Sciences: Perils and Prospects*. Cambridge, UK: Cambridge University Press

Milton J (1667) *Paradise Lost*. Leonard J (ed; 2000). London, UK: Penguin

Nørretranders T (1998) *The User Illusion: cutting Consciousness Down to Size*. New York, NY, USA: Viking

Racine E, Bar-Ilan O, Illes J (2005) fMRI in the public eye. *Nat Rev Neurosci* **6:** 159–164

Robert JS (2004) *Embryology, Epigenesis, and Evolution: Taking Development Seriously*. Cambridge, UK: Cambridge University Press

Rose HA (2000) Consciousness and the limits of neurobiology. In Nunez R, Freeman WJ (eds), *Reclaiming Cognition: the Primacy of Action, Intention and Emotion*. Exeter, UK: Imprint Academic

Rose SPR (2005a) *Lifelines: Biology, Freedom, Determinism*. London, UK: Cape

Rose SPR (2005b) *The 21st Century Brain: Explaining, Mending and Manipulating the Mind*. London, UK: Cape

Rose S (1995) The rise of neurogenetic determinism. *Nature* **373:** 380–382

Rose S (2004) *The Making of Memory*. London, UK: Cape

Rosenberg A (1985) *The Structure of Biological Science*. Cambridge, UK: Cambridge University Press

St Augustine (1961) *Confessions*. [Translated by Pine-Coffin RS.] London, UK: Penguin

Sartre JP (1946) *L'Existentialisme et un Humanisme*. Paris, France: Nagel

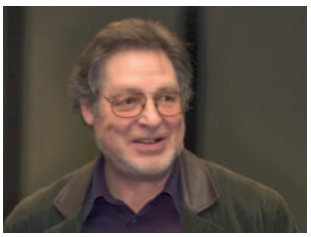
Schacter DL (ed) (1999) *The Cognitive Neuropsychology of False Memories*. East Sussex, UK: Psychology Press

Searle JR (2004) *Mind: A Brief Introduction*. Oxford, UK: Oxford University Press

Sedley S (2004) Responsibility and the law. In Rees D, Rose S (eds), *The New Brain Sciences: Perils and Prospects*. Cambridge, UK: Cambridge University Press

Sperry RW (1985) *Science and Moral Priority: Merging Mind, Brain, and Human Values*. Westport, CT, USA: Greenwood

Zeki S, Goodenough R (2004) Law and the brain: introduction. *Philos Trans R Soc Lond B Biol Sci* **359:** 1661–1665



Steven P.R. Rose is at the Department of Biological Sciences, The Open University, Milton Keynes, UK.
E-mail: s.p.r.rose@open.ac.uk

doi:10.1038/sj.embor.7400566