Paying attention to consciousness

John G. Taylor

Despite being much studied by cognitive neuroscience, consciousness has resisted attempts to understand it. Recent neuroscientific papers on the problem have surprisingly neglected attention as a guide to consciousness. A new neural mechanism is proposed here, guided by a control approach to attention, which identifies the source of consciousness, especially that of the ownership of experience.

Consciousness is a subtle phenomenon, which has so far resisted all attempts to understand it, in spite of the present 'race for consciousness' [1]. Without attention to an input there is no awareness of it [2,3]. Yet several recent papers on consciousness [4–7] have surprisingly neglected attention as a guide to understanding consciousness. To remedy this lacuna, the thesis of this article is that consciousness can be more fruitfully regarded as created by suitably specific processes arising from the movement of attention. This leads to a tentative neural mechanism for the creation of mind from matter.

Attention and the brain

There is now improved understanding of attention through brain imaging and single-cell experiments (reviewed in Box 1). Together these show that attention is controlled from brain regions beyond cortical sensory areas. Attending to an input involves a multiplicative modulation of the relevant input representation in lower sensory areas by feedback from higher regions in parietal and prefrontal cortices. There are still some incomplete features of this story, but in general a control view of attention is now accepted, in which signals from outside 'early' cortex modulate inputs so as to allow selection of a desired target from a set of distracters.

Attention and control

Recent progress in modelling attention based on these results (Box 1) can usefully be viewed from an engineering control framework, allowing a general picture of attention control to be developed that is consistent with the experimental data [8]. Contained within such a control model are the components of primary and secondary sensory and motor cortices, with input activations being controlled by an inverse controller in parietal and frontal sites (indicating where or to what attention should next be directed). The model also has a 'rules module' containing the desired state into which attended input should be transformed, with prefrontal top-down and superior colliculus bottom-up components. A 'forward model', or observer, can also be incorporated, which estimates the present attention state used for rapid error correction before sensory feedback becomes available.

John G. Taylor
Department of Mathematics, King's College, Strand, London, UK WC2R 2LS.
e-mail: john.g.taylor@kcl.ac.uk

Philosophical and meditatory aspects of self

An explanation of consciousness is not in sight from the present attentional viewpoint, even though without attention there is no consciousness of a given input. However, the implicit capture of attention in blindsight subjects [9] and brought about by degraded inputs in normal subjects [10] show that attention is not sufficient for consciousness. We must search within the present model of attention to single out the attentional components of neural activity that would also be sufficient for conscious experience. Recent relevant developments have occurred in the philosophy of consciousness, as it relates to the perception of 'self', especially from phenomenology (see Box 2), and also in understanding the nature of altered states of consciousness (Box 3). In phenomenology there are two components of consciousness: 'consciousness of' (the intentional component) and the pre-reflective self.

The pre-reflective self is experienced as the ownership of one's conscious experience and as the basis of all awareness; without it there would be content, but no owner of that content [11]. There should therefore be room for such ownership information in a model of attention strong enough to contain conscious experience. Coming from an unrelated source, meditation has been claimed by many to lead to the remarkable 'pure consciousness experience' or PCE, which is claimed to be content-free. Recent brain imaging has supported the existence of this distinct state of consciousness (Box 3). Both the pre-reflective self and the PCE are critical components of consciousness.

I have proposed [12] that PCE is a temporal extension of the pre-reflective self [13]. This resolves the contradiction between the Western analytic approach to mind, where consciousness is identified only with content (in contrast to the Western phenomenological approach) and that of Eastern approaches, for which the PCE state is arguably the supreme target of meditation [14]: the Eastern mystic attempts to attain a state already experienced, albeit briefly (and possibly unknowingly), by all humans.

An attention-based control model of consciousness (the CODAM model).

I now attempt to fuse the control model of attention (reviewed in Box 1), with the philosophical and Eastern meditation-based understanding of experience (reviewed in Boxes 2 and 3). In the control model, the observer is assumed to contain a buffered copy of the controller signal, which is used to achieve more rapid updating of the movement control signal [15,16]. Such a copy will not be bound in any attention-based manner to the content of consciousness, because that can only be present on feedback from the sensory

 1364-6613/02/$ – see front matter PII: S1364-6613(02)01890-9

Box 1. Recent advances in analysing and modelling attention

Global brain imaging techniques (PET and fMRI) indicate that a different brain network is involved in moving the focus of attention from that involved in the initial processing of an attended input [a,b]. The regions exercising control of attentional shifting are in parietal and prefrontal sites, whereas sites of processing attended inputs are in primary and secondary unimodal sensory cortices (in the motor cortex for motor responses, for example). When damaged, the parietal and prefrontal sites cause deficits in the speed of attentional shifting [c]. Attention modulates activity in the input sites, as shown both globally using fMRI in humans [d], and by analysis of single-cell recordings in monkey visual cortex [e]. Attention has been shown to affect earlier cortical sites by recurrent feedback, which in anaesthetized monkeys changes the properties of classical receptive fields of visual cortical cells [f]. Detailed timing analyses in humans, using EEG and fMRI, support the existence of attention-controlled feedback [g], as well as control arising from superior parietal sites in the fast dorsal steam, which gates the slower object representations in the ventral stream [h].

Attention control has been found to arise by two mechanisms: via bottom-up signals from unexpected and strong inputs (e.g. a brief flash of light), or by top-down control related to a required goal (e.g. searching for the face of a friend in a crowd). Top-down control is usually in charge; involuntary attentional capture by distracting inputs occurs only if the distractor possesses a similar feature to that of the searched-for target [i]. The lack of attentional capture of transient inputs has been carefully investigated, as has the phenomenon of inattentional blindness, in which apparently important and unexpected events do not draw our attention (or awareness), often to the subsequent surprise of the subjects [j,k].

Psychological models are descriptive, but are based on a process of competition taking place on a high-level module to guide movement of the focus of attention on a lower-level one [l–n]. Neural-network simulations of attention tasks implement this general idea. Simulation of visual search times for targets in the presence of distracters has been performed in a model comprising an input module (representing early visual cortex), a higher-order module in which competition determines where attention is focused (as in the parietal cortex), and an object-coding module that represent objects (as in the temporal cortex).

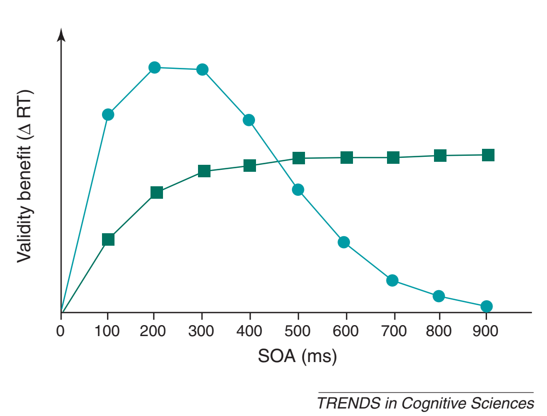

Fig. I. Results of a simulation of the validity benefit (reaction-time difference to invalid cues as compared with valid cues) in the Posner benefit paradigm, as a function of stimulus-onset asynchrony (SOA). Circles: data from exogenous cues; Squares: data from endogenous cues. These simulation results are based on those from a neural-network control model of the movement of attention [s]. The overall shapes (rise and fall times) agree well with human and monkey data.

The experimentally observed linear increase in search time with number of distracters has been replicated in such simulations [o, p]. The template biasing of the competition for attention has also been studied in a model containing explicit frontal sites [q].

An explicit engineering-control framework which fuses these approaches has been outlined [r]. It uses a controlled site (identified as early cortex and temporal lobe), an inverse control module (identified as in parietal lobe), a rules module (in prefrontal cortex) and an observer or forward model (with components in both parietal and prefrontal lobes). This model has close agreement with the dependence of reaction-time decrease achieved by attention to a target (see Fig. I) [s]. Detailed contributions of some of the control components are still being assessed.

References

a Hopfinger, J.B. *et al.* (2000) The neural mechanisms of top-down attentional control. *Nat. Neurosci.* 3, 284–291
b Kastner, S. and Ungerleider, L.G. (2000) Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23, 315–341
c Posner, M. and Petersen, S.E. (1990) The attention system of the human brain. *Annu. Rev. Neurosci.* 13, 25–42
d Friston, K.J. *et al.* (1995) Characterizing modulatory interactions between areas V1 and V2 in human cortex: a new treatment of functional MRI data. *Hum. Brain Mapp.* 2, 211–234
e Reynolds, J.H. *et al.* (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753
f Lamme, V.F. and Roelfsema, K. (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579
g Martinez, A. *et al.* (2001) Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vis. Res.* 41, 1437–1457
h Vidyasagar, T.R. (1999) A neuronal model of attentional spotlight: parietal guiding the temporal. *Brain Res. Rev.* 30, 66–76
i Pashler, H. *et al.* (2001) Attention and performance. *Annu. Rev. Psychol.* 52, 629–51
j Mack, A. and Rock, I. (1998) Inattentional blindness: perception without attention. In *Visual Attention* (Wright, R.D., ed.), pp. 55–76, MIT Press
k Simons, D.J. (2000) Attentional capture and inattentional blindness. *Trends Cogn. Sci.* 4, 147–155
l Posner, M.I. *et al.* (1987) Isolating attentional systems: a cognitive–anatomical analysis. *Psychobiology* 15, 107–121
m LaBerge, D. and Brown, V. (1989) Theory of attentional operations in shape identification. *Psychol. Rev.* 96, 101–124
n Desimone, R. and Duncan, J. (1995) Neural mechanics of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222
o Mozer, M.C. and Sitton, M. (1999) Computational modeling of spatial attention. In *Attention* (Pashler, H., ed.), pp. 341–393, Taylor & Francis
p Deco, G. (1998) Biased competition mechanisms for visual attention in a multimodular neurodynamic system. In *Emergent Neural Computational Architectures Based on Neuroscience* (Wermter, S. *et al.*, eds), pp. 114–126, Springer
q Jackson, S.R. *et al.* (1994) Networks of anatomical areas controlling visuospatial attention. *Neural Netw.* 7, 925–944
r Taylor, J.G. (2001) Attention as a neural control system. In *Proc. Int. Joint Conf. Neural Netw.*, pp. 272–276, IEEE Press
s Taylor, J.G. and Rogers, M. A control model of attention. *Neural Netw.* (in press)

Box 2. New views on the self

Philosophers and psychologists are converging from a variety of angles on the importance of the component of self that is most primitive. Some call it the 'minimal self' [a], others the 'pre-reflective self' [b,c], still others 'pre-reflective self consciousness' [d]. It was missed by the philosopher David Hume when he looked into his own experience and found only 'bundles of sensations'. These sensations of the reflective self have been taken on board by recent Western cognitive analyses of consciousness, which have equated consciousness with 'consciousness of', or intentionality. However, the tradition of phenomenology in Western philosophy, associated with the names of Husserl, Sartre and others, has claimed that there is a primitive sense of self prior to the more complex self experienced as object by introspection.

The primitive self is evident when, for example, one experiences pain. There is no perceptual act that I perform to experience the pain. I do so immediately. This is the source of the 'transparency' of consciousness noted by some philosophers [e]. It is also related to the 'immunity to error through misidentification of the first person pronoun' discussed in [a]. If you say to me '*I am in pain*' it is not sensible for me to ask you '*are you sure it is you who are in pain*?'. Your ownership of the pain is immediate and not something that makes you stop and think '*Is it really me in pain*?'.

'Ipseity' (the intimate sense of ownership that is immune to error through misidentification) is not supposed to arise by some subtle process of self-reflection. So how then can it relate to external input, and thus to the intentional aspect of consciousness?

This quandary has led to many proposed solutions: do away with ipseity altogether [f], do away with consciousness as we experience it and make it a 'center of narrative gravity' [g], make ipseity have mysterious powers (non-material, for example), and so forth. The present consensus among philosophers on the existence of ipseity causes us to face up to the presence of separate contentful and content-free components of consciousness. Western philosophy is presently not able to achieve a unification of these components. Can Eastern ideas help? The phenomenon of 'pure consciousness' has been recognized by some Western researchers [h,i] as providing a missing component of consciousness. It is that of content-free awareness, as if pure consciousness were conscious only of itself. It is then natural to identify this pure conscious experience (PCE) as one in which ipseity has taken over, leading to inhibition of all content. From the attentional point of view, this is an important dissociation of consciousness, into contentful and content-free components.

References

a Gallagher, S. (2000) Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* 4, 14–21

b Parnas, J. (2000) The self and intentionality in the pre-psychotic stages of schizophrenia. In *Exploring the Self* (Zahavi, D., ed.), pp. 115–147, John Benjamins

c Zahavi, D. (1999) Self and consciousness. In *Exploring the Self* (Zahavi, D., ed.), pp. 55–76, John Benjamins

d Sartre, J-P. (2001) *Being and Nothingness* (transl. H.E. Barnes), Routledge

e Metzinger, T. (1995) The problem of consciousness. In *Conscious Experience* (Metzinger, T., ed.), pp. 3–40, Academic Press

f Atkinson, A.P. *et al.* (2000) Consciousness: mapping the theoretical landscape. *Trends Cogn. Sci.* 4, 372–382

g Dennett, D. (2001) Are we explaining consciousness yet? *Cognition* 79, 221–237

h Forman, R.K.C. (1999) What does mysticism have to teach us about consciousness? In *Models of the Self* (Gallagher, S. and Shear, J., eds), pp. 361–377, Academic Press

i Austin, J.H. (1998) *Zen and the Brain*, MIT Press

cortex. The corollary discharge signal will therefore have no content. It can, however, be identified with the experience of 'ownership' – that of the about-to-appear amplified input being attended to. The input activity is amplified by the new attention signal, so accessing its buffer and thereby leading to conscious awareness of the input. The copy of the attention movement signal (the signal that causes the focus of attention on lower modules to be changed) contains the information that this is about to happen, so carries ownership of the consciousness of the input. Such a signal can also grant immunity to error through misidentification [12,15], if the corollary discharge buffer can only let into the working-memory buffer what it has been told to by the 'inverse' attention controller (which transforms current input into desired input). As such, it inhibits all other possible entrants to contentful consciousness in the working memory buffer. This inhibition lasts for a brief period (100–200 ms; [17]) before the attentionally amplified input from sensory cortex arrives. The corollary discharge is then inhibited in its turn. Such complex processing is helped by siting the neural structures for attentional control nearby, in the parietal lobe, which has been identified as crucial for consciousness [18]. A possible control model for this is shown in Fig. 1, which also incorporates an additional observer component.

This results in the CODAM (COrollary Discharge of Attention Movement) model of consciousness [12,15]. According to this model: the pre-reflective self is identified with, and experienced as, the corollary discharge of the attention movement control signal residing briefly in its buffer until the associated attended input activation arrives in its buffer.

The crucial feature is that the buffered corollary discharge signal is *identified* with pre-reflective consciousness. This is a stronger claim than a functionalist one – that pre-reflective consciousness is thereby 'generated' – because there would still be unknown steps in such generation. It is supported by the immunity to error and by PCE experiences, but still needs to be justified by many more features identifiable with other aspects of experience [18].

Attention and the pure conscious experience

The PCE state would be created in CODAM by development, through meditation, of the ability to direct one's attention solely to one's own movement of attention; all content arriving from earlier sensory cortices is inhibited from gaining access to its working-memory buffer. This 'inner eye' of the mind looking at itself has been suggested previously (although unsuccessfully) as a mechanism for the

Box 3. New experiments on meditation

Owing to their subjective nature, meditatory states are difficult to assess. An extreme view could hold that they are 'all in the mind', with no real physical basis. Or perhaps they are simply like falling asleep. Measurements of physiological processes, however, have shown that there are important physical components to meditatory states. More recently, non-invasive brain recording techniques (especially EEG, PET and fMRI) have shown that unique brain states occur in meditation; meditators are not just dropping asleep. The conclusion from millennia of meditation is that the most highly desirable state is that termed 'nirvana', 'samadhi', enlightment or 'samatha', or just the 'pure conscious experience' (PCE). This is claimed to have no content, with consciousness apprehending only itself [a,b].

Early measurements on subjects in meditatory states showed that there were radical alterations in heart rate and oxygen uptake, and a slowing of related bodily processes [c,d]. This was determined as being caused by a decrease of skeletal muscle activation, controlled by chemical neuromodulators released by the brain. With increasing numbers of Westerners practising meditation over the past decade or so, the supply of meditatory adepts for experiments has increased accordingly.

Observations of the nature of brain activity using EEG [c–e] found that under transcendental meditation (TM) subjects had increased low-frequency brain activity, especially the alpha waves (8–9 Hz) in central and frontal regions. Beta activity (12–14 Hz) remained constant or decreased, but fast beta spindles occurred in the deepest meditatory states, most predominantly in anterior EEG channels. This could be related to increased attentional control by frontal brain regions. Increased EEG coherence was also observed, especially in alpha bands. This finding in particular was correlated with subjective reports of the experience of pure consciousness. This interpretation is supported by a more recent study of 20 TM practitioners, who were found to develop increased alpha coherence between frontal and centro-parietal sites [f].

Other recent studies have used PET and fMRI approaches, which are able to probe brain localization more precisely than EEG. A PET study found increased frontal blood flow during meditation compared with rest, with a larger decrease in blood flow in primary and secondary visual regions [g]. An fMRI study found similar frontal, but also parietal, increases in activity, as well as reductions of activity in early visual areas [h].

All these results support the presence of increased prefrontal and parietal activity as part of attention control. There is concomitant reduced visual activity, which corresponds to lowered experiential content of the subjects. However, increased vigilance occurred during the meditatory state, in terms of autonomic responses to sudden stimuli. These physiological studies support the claim that PCE is a distinct state of consciousness, in which attention is attending only to itself.

References

a Wallace, B.A. (1999) The buddist tradition of samatha: method for refining and examining consciousness. *J. Conscious. Stud.* 6, 175–187

b Griffiths, P.J. (1990) Pure consciousness and Indian buddhism. In *The Problem of Pure Consciousness* (Forman, R.K.C., ed.), pp. 71–120, Oxford University Press

c Shear, J. and Jevning, R. (1999) Pure consciousness: scientific exploration of meditation techniques. *J. Conscious. Stud.* 6, 189–209

d Jevning, R. *et al.* (1992) The physiology of meditation: a review. A wakeful hypometabolic integrated response. *Neurosci. Biobehav. Rev.* 16, 415–424

e D'Aquilli, E.G. and Newburg, A.B. (1993) Religious and mystical states. *Zygon* 28, 177–200

f Travis, F. and Wallace, R.K. (1999) Autonomic and EEG patterns during eyes-closed rest and transcendental meditation (TM) practice: the basis for a neural model of TM practice. *Conscious. Cogn.* 8, 302–318

g Herzog, H. *et al.* (1990) Changed pattern of regional glucose metabolism during yoga meditative relaxation. *Neuropsychology* 91, 182–187

h Baerentsen, K.B. *et al.* (2001) Onset of meditation explored with fMRI. *NeuroImage* 13, S297

creation of consciousness [11]; however, it was not based on attention. In CODAM, such a process is given, for the first time, a specific attention-based neural circuitry that allows it to be achieved in the brain without the difficulties of an infinite regress [19]. The key is to use corollary discharge, which allows 'splitting of the attention beam', so that one component can attend to its own copy. Such 'inner attention' also needs to have an associated goal state – that of not possessing any content in experience – set up in the 'rules module', which is located in frontal cortex. Frontal activity has indeed been observed during meditative states, so supports the CODAM model [20]. The model thus provides specific and testable predictions of temporal neural activity at the appropriate sites.

The CODAM model also explains how 'a good proportion of perception occurs without awareness' (Ref. [2], p. 149), such as is seen in the phenomenon of attentional capture by a cue below the threshold for awareness [10]. To explain this phenomenon in terms of CODAM, a processed but degraded cue accesses the inverse controller module and causes attention to be

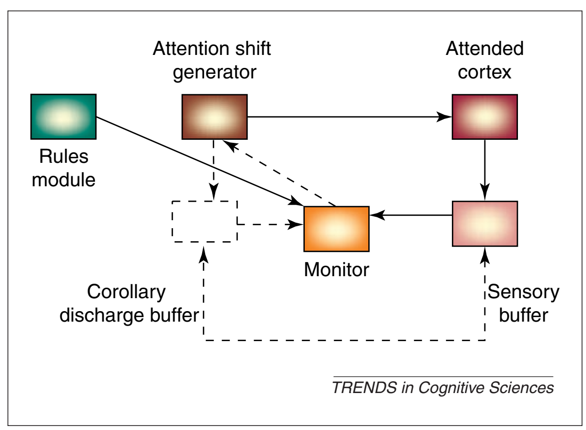

Fig. 1. A simple model for attentional control, which includes an attended site ('early' sensory cortex), an attention shift generator (parietal lobe), a rules module (prefrontal cortex), and a monitor (cingulate cortex), whose activity is based on the error between the required attention state and that presently occurring as determined by a sensory buffer (working-memory buffer). An additional observer component, making use of corollary discharge signal, can also be incorporated (shown in dotted lines). This would speed up the shift of attention (after going through the monitor), and prevent incorrect updating of the sensory buffer by inhibiting it until attention has been moved to the correct location.

Questions for future research

- Where is the corollary discharge buffer sited and what are the details of its functionality?
- How is attention able to guide learning in a conscious manner [22]?
- What is the detailed interaction between parietal lobe and prefrontal cortex that results in an automatic response [22]?
- What is the multi-modal version of the attention-control model?
- How does the CODAM model of consciousness relate to other mental states (NREM sleep, REM sleep, drug-induced states, schizophrenia, the 'dual mystical state', and so on)?

directed to possible inputs. However, because it is degraded, the cue would not access the buffer site of conscious content (thereby losing the competition against the associated corollary discharge), so would not be in awareness. Other cases of implicit attentional capture [2] can be explained along similar lines, as can the attention-catching but unconscious degraded signals in blindsight [9].

Conclusions

An overview has been given of recent developments in modelling attention and understanding the physical correlates of meditatory states. A control model of the movement of attention, supported by empirical data, has been extended to a mechanism for the creation of consciousness, through the CODAM model. This leads to a tentative computational understanding of the minimal or pre-reflective self. The neuroscientific underpinning of certain meditatory states, claimed by some to be the basis of consciousness itself, can begin to be explained. The broad outline given here needs to be filled in with considerable detail (see, for example, Questions for future research), but if the approach is correct, it provides a mechanism to help resolve the 'hard problem' of human consciousness [21].

References

1 Taylor, J.G. (1999) *The Race for Consciousness*, MIT Press
2 Simons, D.J. (2000) Attentional capture and inattentional blindness. *Trends Cogn. Sci.* 4, 147–155
3 Mack, A. and Rock, I. (1998) *Inattentional Blindness*, MIT Press
4 Cotterill, R.M.J. (2001) Cooperation of basal ganglia, cerebellum, sensory cerebrum and hippocampus: possible implications for cognition, consciousness, intelligence and creativity. *Prog. Neurobiol.* 64, 1–33
5 Gallagher, S. (2000) Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* 4, 14–21
6 Frith, C. (1992) *The Cognitive Neuropsychology of Schizophrenia*, Erlbaum
7 Tassi, P. and Muzei, A. (2001) Defining the state of consciousness. *Neurosci. Biobehav. Rev.* 25, 175–191
8 Taylor, J.G. (2001) Attention as a neural control system. In *Proc. Int. Joint Conf. Neural Netw.*, pp. 272–276, IEEE Press
9 Kentridge, B.W. *et al.* (1999) Attention without awareness in blindsight. *Proc. R. Soc. Lond. B Biol. Sci.* 266, 1805–1811
10 McCormick, P.A. (1997) Orienting attention without awareness. *J. Exp. Psychol.* 23, 168–180
11 Zahavi, D. and Parnas, J. (1998) Phenomenal consciousness and self-awareness: a phenomenological critique of representational theory. *J. Conscious. Stud.* 5, 687–705
12 Taylor, J.G. Neural theories of consciousness. In *Handbook of Brain Theory and Neural Computation* (Arbib, M., ed.) MIT Press (in press)
13 Strawson, G. (1999) The self. In *Models of the Self* (Gallagher, S. and Shear, J., eds), pp. 1–24, Academic Press
14 Forman, R.K.C. (1990) *The Problem of Pure Consciousness*, Oxford University Press
15 Taylor, J.G. (2000) The central representation: the where, what, and how of consciousness. In *The Emergence of Mind* (White, K.E., ed.), pp. 149–170, Fondazione Carlo Erba
16 Desmurget, M. and Grafton, S. (2000) Forward modeling allows feedback control for fast reaching movements. *Trends Cogn. Sci.* 4, 423–431
17 Reynolds, J.H. *et al.* (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753
18 Taylor, J.G. (2001) The importance of the parietal lobes for consciousness. *Conscious. Cogn.* 10, 379–417
19 Shoemaker, S. (1868) Self-reference and self awareness. *J. Philos.* 65, 556–579
20 Lou, H.C. *et al.* (1998) A ^{15}O-H_2O PET study of meditation and the resting state of normal consciousness. *Hum. Brain Mapp.* 7, 98–105
21 Chalmers, D. (1996) *The Conscious Mind: Towards a Fundamental Theory*, Oxford University Press
22 Willingham, D. (1998) A neurophysiological theory of motor skill learning. *Psychol. Rev.* 105, 558–584