Consciousness: mapping the theoretical landscape

Anthony P. Atkinson, Michael S.C. Thomas and Axel Cleeremans

What makes us conscious? Many theories that attempt to answer this question have appeared recently in the context of widespread interest about consciousness in the cognitive neurosciences. Most of these proposals are formulated in terms of the information processing conducted by the brain. In this overview, we survey and contrast these models. We first delineate several notions of consciousness, addressing what it is that the various models are attempting to explain. Next, we describe a conceptual landscape that addresses how the theories attempt to explain consciousness. We then situate each of several representative models in this landscape and indicate which aspect of consciousness they try to explain. We conclude that the search for the neural correlates of consciousness should be usefully complemented by a search for the computational correlates of consciousness.

A.P. Atkinson is at The Psychology Dept, King Alfred's College, Sparkford Road, Winchester, Hampshire, UK SO22 4NR.

tel: +44 1962 827455
fax: +44 1962 827437
e-mail: A.Atkinson@wkac.ac.uk

M.S.C. Thomas is at The Neurocognitive Development Unit, Institute of Child Health, University College London, London, UK WC1N 1EH.

e-mail: M.Thomas@ich.ucl.ac.uk

A. Cleeremans is at The Cognitive Science Research Unit, Université Libre de Bruxelles CP 122, Avenue F-D. Roosevelt, 50, 1050 Bruxelles, Belgium.

e-mail: Axel.Cleeremans@ulb.ac.be

The search for the 'neural correlates of consciousness' (NCC) is burgeoning[1–3] (see Box 1), thanks in particular to the increased availability of brain imaging techniques. As a recent article points out[3], the search for the NCC will undoubtedly increase our understanding of the neural bases of conscious experience. However, the emergence of methodologies that appeal simultaneously to computational modeling and to empirical exploration of both behavior and brain function would seem to mandate that we go beyond the establishment of correlations and attempt to elaborate detailed theories of consciousness. Many such theories have appeared recently within the cognitive neurosciences. What distinguishes these proposals from earlier ones is an emphasis on information processing, that is, on the notion that cognition involves processes that manipulate representations.

What exactly do such theories try to explain? A straightforward answer is that information-processing theories of consciousness aim to explain how it is that certain items of information in the brain are available to consciousness while others are not. Some theorists have thus proposed that what determines the contents of consciousness is some special way in which information is represented. Others have proposed instead that the crucial factor is a special type of information process. In this paper, our goal is to survey what one could call the search for the 'computational correlates of consciousness' and to organize representative proposals so as to uncover their underlying assumptions. Before we start, however, it is worth considering the different concepts that have been identified as falling under the label of 'consciousness'.

In this article, our goal is to survey what one would call the search for the 'computational correlates of consciousness' and to organize representative proposals so as to uncover their underlying assumptions. Our conceptual landscape is defined by two dimensions. First there is a 'process versus representation' dimension. Is consciousness supposed to arise from particular computations that are performed over representations in the brain, or does it arise from some intrinsic property of the representations themselves? Second there is a 'specialized versus non-specialized' dimension. Is consciousness assumed to involve mechanisms dedicated to consciousness, or is it assumed to arise from the appropriate kinds of computations or representations wherever in the brain these may occur? These two dimensions generate four possible views: (1) consciousness arises from particular computational processes, provided that these occur in the mechanism that is dedicated to consciousness; (2) consciousness arises from particular kinds of representations, provided that these occur in the mechanism that is dedicated to consciousness; (3) consciousness arises from particular processes wherever they may occur; and (4) consciousness arises from particular kinds of representations wherever they may occur. Before we start developing these proposals, however, it is worth considering the different concepts that have been identified as falling under the label 'consciousness'.

What do we mean by 'conscious'?

The concept of consciousness is notoriously difficult to define, first because it refers to heterogeneous phenomena, and second because it is difficult to measure objectively (see Box 2). Yet, the scientific study of consciousness thrives, for

1364-6613/00/$ – see front matter PII: S1364-6613(00)01533-3

Box 1. Correlating consciousness: the search for the neural correlates of consciousness (NCC)

The search for the NCC involves isolating the neural processes that correlate with various states of consciousness, including background states of consciousness (awake, sleeping, dreaming) or, more commonly, specific contents of consciousness. The latter case involves discovering the representational contents of neural systems and determining whether they match up with the contents of consciousness.

Chalmers (Ref. a) suggests a working definition of an NCC:

… [An NCC is] *a minimal neural system such that there is a mapping of states of that system to states of consciousness, where a given state of the neural system is sufficient, under certain conditions, for the corresponding state of consciousness.*

The 'certain conditions' involve normal brain functioning and can include ecologically invalid inputs or limited direct brain stimulation, but perhaps not cases of lesions as these may alter the brain's functional architecture. This definition leaves open the question of whether there are mechanisms dedicated to consciousness, and the question of whether the mechanisms subserving consciousness are localized or distributed across the brain. Moreover, it recognizes that the NCC may be different for different states of consciousness (e.g. background versus content).

Chalmers summarizes a number of proposals that have been put forward concerning the nature and location of the NCC. These include 40 Hz oscillations in the cerebral cortex (Ref. b), intralaminar nuclei in the thalamus (Ref. c), reciprocal signaling in thalamocortical systems (Ref. d), 40 Hz rhythmic activity in thalamocortical systems (Ref. e), an extended reticular–thalamic activation system (Ref. f), neural assemblies bound by *N*-methyl-D-aspartate (NMDA; Ref. g), certain neurochemical levels of activation (Ref. h), certain neurons in the inferior temporal cortex (Ref. i), neurons in the extrastriate visual cortex projecting to prefrontal areas (Ref. j), and visual processing within the ventral stream (Ref. k).

A range of empirical techniques is currently in use to reveal the NCC (Ref. l). These techniques attempt to measure neural activity under conditions that differ in terms of the relation between behavior and awareness. The intention is to isolate the neural activity that always correlates with awareness under the following situations, which relate behavior and awareness:

- demonstrations of appropriate behavior to stimuli in the absence of reported perceptual awareness of those stimuli (e.g. masked priming, implicit learning, blindsight, amnesia, visual form agnosia, 'anarchic' limbs)
- reports of conscious perceptions in the absence of sensation (phantom limbs, hallucinations, false memories)
- a constant stimulus input that generates alternating conscious perceptual states (binocular rivalry, bistable images)
- stimuli that are perceived but ignored (neglect) or not acted towards appropriately (optic ataxia)
- temporal dissociations between intentional movement towards stimuli and awareness of those stimuli
- behavior that is intended consciously but is without any associated behavior ('movement' of phantom limbs, motor imagery).

Three points are of note in this list. First, there is much use of lesion evidence. This needs to be interpreted with care because, as Chalmers points out, the NCC may be different in a lesioned system than in an intact one (Ref. a). Second, the search for the neural correlates of the contents of consciousness places an emphasis on the development of techniques that can establish the contents of neural representations (Ref. a). The most successful current method records the activity of single cells, whereby the receptive and projective fields of neurons can be determined. In other words, we can establish what stimuli a given neuron tends to respond to. We can then explore whether the activity of this cell correlates with reported awareness (see Box 2 for approaches to measuring awareness). For ethical reasons, this technique has been used mostly with monkeys (Refs i,m,n). For human subjects the alternative is brain imaging, but this is currently a much less precise method for determining the contents of neural representations. Third, and last, neural activity can be described at different levels of abstraction. At which level we should expect to find correlations with conscious experience is an open question (see Box 4). It is possible that this level is purely computational, such that there are no NCCs, only computational correlates (in which case the CCCs would, at least in principle, be multiply realizable). However, the wealth of recent empirical evidence suggests correlations may be found at less abstract levels of description.

References

a Chalmers, D.J. What is a neural correlate of consciousness? In *Neural Correlates of Consciousness: Conceptual and Empirical Questions* (Metzinger, T., ed.), MIT Press (in press)
b Crick, F. and Koch, C. (1990) Towards a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275
c Bogen, J.E. (1995) On the neurophysiology of consciousness, part I: An overview. *Conscious. Cognit.* 4, 52–62
d Edelman, G.M. (1989) *The Remembered Present: A Biological Theory of Consciousness*, Basic Books
e Llinas, R.R. *et al.* (1994) Content and context in temporal thalamocortical binding. In *Temporal Coding in the Brain* (Buzsaki, G. *et al.*, eds), Springer Verlag
f Newman, J. and Baars, B.J. (1993) A neural attentional model for access to consciousness: a Global Workspace perspective. *Concepts Neurosci.* 4, 255–290
g Flohr, H. (1995) Sensations and brain processes. *Behav. Brain Res.* 71, 157–161
h Hobson, J.A. (1997) Consciousness as a state-dependent phenomenon. In *Scientific Approaches to Consciousness* (Cohen, J. and Schooler, J., eds), pp. 379–396, Lawrence Erlbaum
i Sheinberg, D.L. and Logothetis, N.K. (1997) The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3408–3413
j Crick, F. and Koch, C. (1995) Are we aware of neural activity in primary visual cortex? *Nature* 375, 121–123
k Milner, A.D. and Goodale, M.A. (1995) *The Visual Brain in Action*, Oxford University Press
l Frith, C. *et al.* (1999) The neural correlates of conscious experience: an experimental framework. *Trends Cognit. Sci.* 3, 105–114
m Logothetis, N.K. and Schall, J. (1989) Neuronal correlates of subjective visual perception. *Science* 245, 761–763
n Leopold, D.A. and Logothetis, N.K. (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379, 549–553

good reasons: progress in science neither requires full analyses of the concepts it purports to study nor foolproof measuring instruments.

Conceptual analysis nevertheless plays an important role in science (it has been very much to the fore in the work of many great physicists, for example). In

Box 2. Measuring consciousness: continuing challenges

Although consciousness originally constituted the primary subject matter of scientific psychology, by the 1940s introspection had become a disreputable method for collecting data in experimental psychology. The subsequent requirement that evidence be objectively verifiable represents a significant problem for the scientific study of consciousness. Not only do we lack a 'consciousness meter' (Ref. a) that could establish what a person (or animal) is conscious of, or whether it is conscious at all, but also, even if we had such a device, it would not solve all the relevant functional measurement problems (Ref. b). Current research is thus predicated on the use of behavioral markers for consciousness. The most commonly used marker is verbal report: we simply ask the person whether they see a light or not, or which view of an ambiguous figure such as the Necker cube they are currently experiencing.

However, perhaps verbal report should not be given such a privileged status. Other behavioral measures might be used, such as communicative gestures or movements, button presses or blinks. Such alternative paradigms are in fact necessary when studying animals because it is not possible for the experimenter to agree with the animal on which communicative acts will indicate awareness. Instead, the animal must be trained to produce a behavior that will stand as a marker for its conscious experience.

To take an example, let us say that we wished to demonstrate blindsight in animals. To do so, we must find stimulus-relevant behavior in response to a visual stimulus, but in the absence of the marker for conscious experience of that stimulus. We begin by defining two tasks. One task stands as a measure of stimulus-relevant behavior, such as a forced-choice categorization (e.g. deciding whether a light is to the left or the right of a vertical line). A second task stands as the marker for consciousness, such as classification of the visual scene (e.g. is the screen blank or is there a light there?). Inconsistency between performance on the two tasks (e.g. the light is reliably categorized as falling to the correct side of the vertical line while the screen is identified as blank) would thus be taken to indicate performance in the absence of awareness. Such an approach allowed Cowey and Stoerig to demonstrate blindsight in monkeys with lesions to the striate cortex of the left hemisphere (Refs c,d).

The challenge associated with using different behavioral markers for consciousness is that they fail to produce congruent results in certain situations, such as in implicit learning or subliminal perception paradigms (Ref. e). These dissociations raise the difficult methodological issue of determining the extent to which they should be taken as indications of the existence of several distinct processing systems, only some of which are associated with awareness, or rather as a result of the varying sensitivity of the corresponding measures to conscious content. These measurement challenges find an echo in the range of theoretical positions that have been expressed about the relationship between conscious and unconscious processing (see Box 5).

Neuroscientific approaches are not immune to these challenges. For instance, Tononi and Edelman's 'dynamic core' hypothesis (Ref. f) proposes a neural marker of consciousness according to which one is conscious of information that is represented in a simultaneously integrated and differentiated shifting pattern of neural activity. This proposal could perhaps be tested by correlating subjects' verbal reports of their current experience with observed patterns of neural activity. Those neural patterns that fulfil the integrated-and-differentiated criterion should correspond to the verbally reported contents of experience. However, in performing such an experiment, one would clearly rely on testing one assumption (about a certain neural marker for consciousness) using another (about a certain behavioral marker for consciousness).

References

a Chalmers, D.J. (1998) On the search for the neural correlate of consciousness. In *Toward a Science of Consciousness II: The 1996 Tucson Discussions and Debates* (Hameroff, S. *et al.*, eds), pp. 219–229, MIT Press

b Cleeremans, A. and Haynes, J-D. (1999) Correlating consciousness: a view from empirical science. *Revue Internationale de Philosophie* 3, 387–420

c Cowey, A. and Stoerig, P. (1995) Blindsight in monkeys. *Nature* 373, 247–249

d Stoerig, P. and Cowey, A. (1997) Blindsight in man and monkey. Invited review article. *Brain* 120, 535–559

e Cleeremans, A. *et al.* (1998) Implicit learning: news from the front. *Trends Cognit. Sci.* 2, 406–416

f Tononi, G. and Edelman, G.M. (1998) Consciousness and complexity. *Science* 282, 1846–1851

our view, the most useful recent attempt to articulate distinct concepts of consciousness is Block's[4,5], who distinguishes four concepts: access consciousness, phenomenal consciousness, monitoring consciousness, and self-consciousness.

(1) Access consciousness (A-consciousness) refers to our ability to report and act on our experiences: to yell in pain upon cutting your finger, to seek to stem the flow of blood and alleviate the pain, to tell someone what you've done and ask for a bandage. More formally, for a person to be in an A-conscious state means that there is a representation in that person's brain, the content of which is available for verbal report and for high-level processes such as conscious judgements, reasoning, and the planning and guiding of action.

(2) Phenomenal consciousness (P-consciousness) refers to the qualitative nature of experience. P-consciousness applies most straightforwardly to bodily sensations and perceptual experiences: there is 'something that it is like' to experience the pain of the throbbing toe or the piercing tones of an alarm clock[6]. In addition, whereas there is nothing that it is like to be a book or a toaster, there is something that it is like to be you or me.

(3) Monitoring consciousness (M-consciousness) refers to thoughts about or awareness of one's sensations and percepts, as distinct from those sensations and percepts themselves.

(4) Finally, self-consciousness (S-consciousness) refers to thoughts about or awareness of oneself.

As Block notes, few existing information-processing theories clearly distinguish between varieties of consciousness, and thus sometimes conflate two or more concepts. Still, it is clear that almost all of these theories of consciousness are accounts of the mechanisms underlying A-consciousness. Many of these theories are also claimed to be accounts of the mechanisms underlying P-consciousness.

Box 3. Cognition and consciousness: Commander Data meets the zombies

In the face of the challenges associated with measuring consciousness (see Box 2), theories of cognition necessarily have to make assumptions about the relationship between conscious and unconscious processing. The notion that information processing can occur outside consciousness has a long and controversial history. Indeed, while it cannot be denied that at least some neural processing occurs outside consciousness, different theories have made widely disparate assumptions about the exact extent to which cognition (learning, perception and memory) and consciousness co-occur.

The two extreme positions are well illustrated by attempts to reconcile classic information-processing models of cognition with the phenomena of implicit cognition (e.g. subliminal perception, implicit learning, blindsight). For what one could dub 'Commander Data' theories, every information-processing event that counts as cognitive is also conscious, whereas for 'Zombie' theories, every information-processing event that counts as cognitive is inherently unconscious and only becomes optionally available to consciousness.

Star Trek's character Data is an android whose bodily and cognitive innards are fully transparent to itself. Except in rare circumstances (which often tend to be described as the result of some sort of dysfunction), Data is thus capable of describing in uncanny detail each and every aspect of its internal states: How much force it is applying when attempting to pry open a steel door, how many circuits are currently active in its positronic brain, etc.

Commander Data theorists likewise assume that cognition is, at least potentially, fully transparent, that is: (1) that whatever knowledge is expressed through behavior is also transparently available to introspection; and (2) that consciousness allows access, with sufficient effort or attention, to all aspects of our inner cognitive lives.

In contrast, the famed philosophical zombies (Ref. a) are perfectly opaque: Whatever knowledge currently influences their behavior can be neither explicit nor conscious because, by definition, they lack conscious experience. Zombie theorists thus take it as a starting point that consciousness has an epiphenomenal character: There is a zombie within you that is capable of processing all the information your conscious self can process consciously, with one crucial difference, 'All is dark inside' (Ref. a); your zombie is unconscious. From this perspective then, cognition is inherently opaque and consciousness, when present, offers but a very incomplete and imperfect perspective on internal states of affairs.

Note that these two perspectives are not intended as philosophical thought experiments but as (slightly caricatured) illustrations of currently held theoretical positions. Echoes of the two perspectives can be found in recent debates about implicit learning and memory. For some authors (Refs b,c), cognition is systematically accompanied by awareness (and hence the phenomena of implicit cognition, such as subliminal perception or implicit learning, should be dismissed as methodological artefacts), while for others (Ref. d), knowledge can be acquired and deployed without necessarily being available to conscious awareness. A challenging conceptual issue in this respect has to do with our assumptions about the nature of representation: Does a pattern of neural activity count as a cognitive representation when it is not conscious? Should one consider that representations only become conscious when one can entertain certain 'higher-order' representations about them (Ref. e)? Are 'first-person' representations different from 'third-person' representations? These issues continue to be the object of lively debate.

References

a Chalmers, D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press

b Shanks, D.R. and St John, M.F. (1994) Characteristics of dissociable learning systems. *Behav. Brain Sci.* 17, 367–395

c O'Brien, G. and Opie, J. (1999) A connectionist theory of phenomenal experience. *Behav. Brain Sci.* 22, 127–196

d Cleeremans, A. (1997) Principles for implicit learning. In *How Implicit is Implicit Learning?* (Berry, D., ed.), pp. 196–234, Oxford University Press

e Dienes, Z. and Perner, J. (1999) A theory of implicit and explicit knowledge. *Behav. Brain Sci.* 22, 735–808

Explanations of monitoring consciousness and of self-consciousness can sometimes be discerned in computational theories, although they tend to play second fiddle to explanations of access and phenomenal consciousness. So, for each of the theories we discuss, we identify whether it is best interpreted as examining A-consciousness, P-consciousness, or both A and P.

Two issues deserve further comment. First, the extent to which different forms of consciousness can actually be dissociated remains unclear. For instance, can information be available for the control of action yet not be phenomenally conscious? Different models make different assumptions regarding such issues (see Box 3 for further discussion).

Second, it remains uncertain whether scientific accounts of P-consciousness are possible. This is the 'hard problem'[7] or the problem of the 'explanatory gap'[8,9], that is, the gap in our understanding of how it is that physiological and information-processing events in our brains can be responsible for the 'what it is like'-ness, the experiential or phenomenal aspect of consciousness. It is the question of 'why' experiences 'feel like anything at all'. It is because of this explanatory gap that current scientific theories of P-consciousness seem bound to be incomplete. We remain agnostic about the possibility of a scientific account of phenomenal consciousness. Perhaps our increasing knowledge of the neural and computational correlates of consciousness will eventually bridge the explanatory gap, or perhaps that knowledge will change the way we conceive P-consciousness such that the explanatory gap disappears[10–12].

Mapping the conceptual landscape

Let us now explore the properties of computational models of consciousness. How do such models explain one or more of the different aspects of consciousness, and which information-processing principles do the models deploy? We suggest that existing computational models of consciousness all appeal to properties that can be grouped along two principal dimensions, with differing degrees of commitment along each dimension:

- The process versus representation dimension opposes models that explain consciousness in terms of specific processes operating over mental representations, with models that explain consciousness in terms of intrinsic properties of mental representations.
- The specialized versus non-specialized dimension contrasts models that posit information-processing systems dedicated to consciousness with models for which consciousness can be associated with any information-processing system as long as this system has the relevant properties.

Processing or representation?

Process theories assume that consciousness depends on certain functional or relational properties of representational vehicles, namely, the computations in which those vehicles engage. On this view, representational contents are conscious when their vehicles have some privileged computational status, independently of any particular intrinsic property of those vehicles. What counts is 'what representational vehicles do, rather than what they are' (Ref. 13, p. 128).

For vehicle theories, on the other hand, consciousness is determined by intrinsic properties of representational vehicles, independently of any computations in which those vehicles engage. On this view, consciousness, and particularly P-consciousness, should be explained in terms of the way information is represented[13]. Thus, when you observe the scene outside your window, what it is like for you to have that visual experience is to be explained in terms of the way in which the objects in your visual field – their shape, color, relative size and location – are represented in your brain. In other words, consciousness depends on some aspect of the physical medium of representation[14]. For instance, O'Brien and Opie's view is that conscious information is explicitly represented information, by which they mean that it is encoded 'in such a way that each distinct item of data is encoded by a physically discrete object' (Ref. 13, p. 128). Various other properties of representations have also been proposed, such as their stability, their strength, or their distinctiveness[15–17].

Process theories might seem more appropriate for accounts of A-consciousness than vehicle theories, as A-consciousness is typically explained in terms of relations amongst information-processing systems[4]. By contrast, P-consciousness is often characterized as an intrinsic property[4], which might seem more consistent with vehicle theories but, as we shall see, there are both vehicle and process theories of A- and P-consciousness.

Process and representation are clearly not independent in any computational theory. A process must operate over a given set of representations (though see Ref. 18). Yet the emphasis here concerns the particular property of the computational system that is associated with consciousness of a given sort.

Specialized machinery for consciousness?

Our second dimension is rooted in a standard assumption in cognitive science, namely that the mind, viewed as a complex information-processing machine, can be decomposed into functional modules, each specialized for a certain set of tasks (e.g. see Refs 19 and 20). For instance, different aspects of object perception, face perception and language processing are thought to be subserved by specialized or dedicated modules (e.g. Refs 21 and 22), as are some highly complex cognitive capacities, such as reasoning, moral belief and psychological understanding (e.g. Refs 23 and 24). This being the case, it is therefore tempting to ask whether there might be a 'consciousness module'.

Claims about the existence of such a module have tended to assume both that a specialized system subserves consciousness and that the information represented in this subsystem is information of which the owner of the brain is conscious. Thus, according to such models, we are conscious of the fact that an object has a certain shape, or that it moves in a particular way, by virtue of having information about the object's shape or movement represented in the 'consciousness module'. Such a module is therefore an information-processing mechanism specialized for producing conscious experience, and whose representational contents are the contents of that experience. Candidate consciousness modules have included temporary memory systems and central executive systems[25–27]. In most theories, such central systems play the functional role of making their informational contents globally available, that is, available (although perhaps not actually used) for verbal report and for high-level processes such as those underlying conscious judgements, reasoning and the planning and guiding of action[28] (for a review see Ref. 1). The concept of consciousness in play here is A-consciousness.

Not all models make the assumption that consciousness derives from the operations of a specialized central module. How are the contents of consciousness to be explained in these alternative accounts? Two possibilities can be distinguished. First, the contents of experience might be the representational contents of some neural network – but not one specialized for producing that experience. For instance, the contents of visual experience would be the representational contents of certain modules within the visual system. Alternatively, the contents of experience might not be the contents of any modules at all but rather emerge from the collective activity of many components distributed both spatially and functionally across the brain, none of them responsible for consciousness on its own. From this perspective, consciousness would be a property of the whole system.

Locating theories on the map

We now explore how different existing computational models of consciousness can be organized along the two dimensions outlined in the previous section. Figure 1 illustrates where we believe 12 representative theories of consciousness lie in the space that results from crossing our two dimensions. As can be seen, the bulk of the models fall into the 'specialized process' quadrant of our map. In the following, we describe a paradigmatic example from each quadrant, and illustrate why it lands where it does on the landscape (see Ref. 29 for details about more of the models). Our intention here is not to evaluate the respective claims of each model but to illustrate the various assumptions that have been in play in the development of computational theories of consciousness.

Specialized vehicle theories

We start with a model that is almost 30 years old, and a lonely model at that, because it is the only one falling within the 'specialized vehicle' quadrant of our map. Even though Atkinson and Shiffrin[25] did not frame their short-term memory model as an explicit theory of consciousness, we can nevertheless view this model as a specialized vehicle theory because it assumes: (1) that the contents of conscious experience are the representational contents of a short-term memory store; and (2) that to be a content of this short-term store, representations must be above a certain level of 'activation'.

In Atkinson and Shiffrin's model, memory is divided into a long-term store (LTS) and a short-term store (STS). The contents of the STS are all and only those representations with a state of activation above a certain threshold. Whether a representation reaches this critical threshold and maintains its activation above that level is determined by various STS control processes initiated and maintained by the subject, such as the rehearsal of phonological information and the imaging of visual information. Crucially, the model assumes that the informational contents of the STS constitute the contents of consciousness, i.e. 'the thoughts and information of which we are currently aware' (Ref. 25, p. 83). It would seem that a theory of both A- and P-consciousness is in the offing.

Other more recent theories of consciousness are likewise based on assumptions about short-term memory but they generally tend to be specialized process theories rather than specialized vehicle theories (e.g. Ref. 26). That is, these models all tend to assume that a representation is conscious not because of some property of its vehicle (e.g. whether it is above a certain level of activation) but because of its functional role (e.g. featuring in the processing activities of an executive component of a working memory system).

Specialized process theories

Specialized process theories assume that consciousness arises from specific computations in a dedicated mechanism. A paradigmatic example is Schacter's model[27], which was proposed as an explanation of neuropsychological disorders that demonstrate dissociations of 'covert' abilities from 'overt' abilities (e.g. covert from overt face recognition). In this model, the outputs of domain-specific perceptual and memory modules can form the inputs to a 'conscious awareness system' (CAS). Cases of preserved covert abilities in the absence of the corresponding overt abilities are explained by the model as the result of damage to the connections between an intact domain-specific module and the intact CAS.

The CAS serves three functions: (1) it integrates various domain-specific outputs; (2) it sends this integrated information to the 'executive system' (ES); and most importantly, (3) its operation allows 'phenomenal awareness of on-going mental activity' (Ref. 27, p. 369). In other words, our ability to enjoy the phenomenal experiences associated with, say, words that we hear or previously learned facts results from the proper functioning of the CAS. Because the contents of such experiences are the informational contents of the CAS, it can be thought of as a consciousness module. Although Schacter does not identify a particular location for the CAS, he nevertheless cites evidence indicating the importance of a posterior portion of cortex, critically involving the inferior parietal lobes, in perceptual awareness.

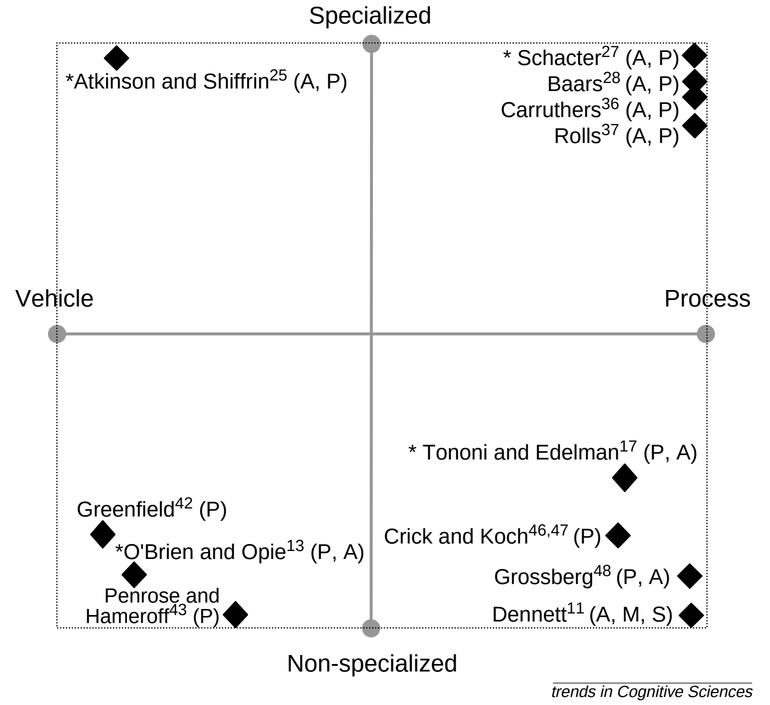


Fig. 1. A 2-D map of computational theories of consciousness. This represents our best guess at where a small selection of theories lie with respect to our two core dimensions. The distance that each theory is placed along each dimension represents how strongly the theory is committed to this feature as its core proposal for an explanation of consciousness. Letters in parentheses indicate the variety of consciousness that each theory is primarily aimed at explaining (phenomenal, access, monitoring and self consciousness). Where there is more than one letter, the first letter is the primary aspect. Theories marked with an asterisk are the paradigmatic examples discussed in the text.

The CAS is taken to underpin P-consciousness, but the model also provides an account of A-consciousness. A-consciousness depends on both the CAS and the ES: the CAS makes information available for further processing, while the ES regulates attention and voluntary activities, functions associated with the frontal lobes[30–32].

Schacter's model is a paradigmatic process theory of consciousness because it proposes that the contents of consciousness are those representational contents available to enter into the processing activities of the ES, which is the mechanism underpinning complex, flexible reasoning and rationally guided action (i.e. the mechanism underpinning A-consciousness).

Three other models can likewise be described as specialized process theories. Baars[28] proposes a mechanism called the 'global workspace' (GW) as the computational substrate of consciousness. [It is the computational substrate for both A- and P-consciousness, as for Baars, P = A (Ref. 33).] According to this proposal, information is conscious by virtue of being represented in the GW and is thus available to the processes underlying rationally guided actions. Baars and his colleagues[34,35] have accumulated considerable evidence suggesting that, for conscious perceptual experience, the neural substrate of the GW is constituted by feedback loops between the thalamus and cortical primary sensory projection areas, with global control 'being effected via cortical "gating" of a strategic thalamic nucleus' (Ref. 35, p. 1195). Another popular candidate for a process that can underpin consciousness is one in which the objects of representations (what they are about) are other representations. Such processes of higher-order representation may require dedicated mechanisms, as in the theories of Carruthers[36] and Rolls[37], which propose that a mental state is conscious in virtue of the subject having a higher-order thought about that state[38,39]. Rolls more than Carruthers is concerned to find the neural substrate of the ability to form higher-order thoughts: he suggests that the language systems and their connections to other cortical areas are critical for this ability.

Reflecting an increasing dissatisfaction with classic information-processing approaches to cognition, a recent trend in the study of consciousness has been to refrain from positing any specialized mechanisms for consciousness. We now consider two such theories: one a non-specialized vehicle theory; the other a non-specialized process theory.

Non-specialized vehicle theories

Non-specialized vehicle theories of consciousness assume that it depends only on specific properties of representations. For instance, in O'Brien and Opie's[13] 'connectionist theory of phenomenal experience', stable patterns of activation are the vehicles of explicit representation in parallel distributed processing (PDP) networks. When realized in neural networks, a stable pattern of activation is achieved when the 'constituent neurons are firing simultaneously at a constant rate' (Ref. 13, p. 139). Assuming that all explicit representations are conscious representations, they then claim that 'phenomenal experience is identical to the brain's explicit representation of information, in the form of stable patterns of activation in neurally realized PDP networks' (Ref. 13, p. 138); but see Refs 40 and 41.

O'Brien and Opie claim their theory is a vehicle theory of P-consciousness because it proposes that phenomenal experience emerges whenever the representations exhibit a given intrinsic property (stability), whatever their functional roles. The theory is also non-specialized, as it makes no commitments about phenomenal experience being generated by specialized components. Indeed, according to O'Brien and Opie's theory, many explicitly tokened representations can exist simultaneously in various parts of the brain, each of which will generate consciousness: P-conscious states, in this view, are complex, multimodal aggregates of distinct phenomenal contents (Ref. 13, pp. 140–142).

While O'Brien and Opie set their sights on an account of P-consciousness, they nevertheless suggest that the notion of stability may also provide an account of A-consciousness, as 'only stable patterns of activation can facilitate meaningful communication between PDP networks, and hence contribute to coherent schemes of action' (Ref. 13, p.140). Although this is a process account of A-consciousness, in their view it is nevertheless contingent on P-consciousness, for which they give a vehicle theory. The stable representation is conscious information; it is secondarily more useful in facilitating inter-network communications.

Two other models that consider that consciousness depends on intrinsic properties of representations are Greenfield's[42], and Penrose and Hameroff's[43]. Greenfield proposes that the degree of consciousness (defined as a continuum) corresponds to sheer amount of neural firing in cortex, as one representation comes to dominate over others in a competitive process. Penrose and Hameroff's controversial proposal is that consciousness corresponds to quantum events occurring in tiny protein structures within neurons (see Ref. 44 for an appraisal).

Non-specialized process theories

Non-specialized process theories of consciousness assume that consciousness depends on specific processes that can take place in any region of the brain. A recent proposal that makes this assumption explicit is Tononi and Edelman's 'dynamic core hypothesis'[18], according to which the signature of the neural processes underpinning consciousness are strong, stable patterns of reciprocal signaling (the 'core') that integrate information from amongst a continually changing selection of numerous widely distributed brain areas (see also Ref. 45). The theory is a process theory in the sense that conscious experience is seen as being dependent on certain neural processes (those resulting from reciprocal signaling) rather than on some inherent property of certain neural representations. It is also a nonspecialized theory, as it does not identify any particular mechanism or set of mechanisms specialized for producing consciousness. Admittedly, Tononi and Edelman do see the thalamocortical areas as playing a crucial role in many types of conscious state, especially in sensory–perceptual awareness, but this is not to claim that these areas are functionally specialized solely, or even mainly, for producing consciousness.

While Tononi and Edelman do not distinguish different concepts of consciousness, their proposal is nevertheless readily interpreted as suggesting that changes in P-consciousness accompany changes in A-consciousness. According to their account, the strong and stable patterns of neural activity constituted by reciprocal signaling underlie coherent, multimodal phenomenal experiences: different properties of the conscious scene are bound together by virtue of the reciprocal signaling between neuronal groups that code for those properties. The implication is that at a given moment the information processed by the neural maps that contribute to these strong and stable patterns of activity form the contents of consciousness. This is the information that subjects are able to report, to use as premises in their reasoning, and to guide their intentional actions (i.e. information that is A-conscious). We also place in this quadrant the theories of:

- Dennett[11], for whom conscious representations are those that drive the behavior taken as a measure of consciousness (see Box 3)
- Crick and Koch[46,47], for whom synchronous firing of neurons in visual cortex might solve the binding

Box 4. Binding and consciousness: can a neuron give rise to consciousness when it is not firing?

Process theories assume that the contents of consciousness correspond to the results of certain special computations carried out in neural structures. Representational codes, whether digital or analog, treat a reduction of activity among elements as informative (i.e. the zeroes count as much as the ones). Thus when a representation forms, neurons that are not firing play as much of a role in generating the phenomenal state as neurons with elevated firing rates. Almost all theorists consider that neurons may also fire without contributing to consciousness if they do not take part in relevant computations or carry the intrinsic property required of a vehicle theory. Taken together, these two ideas suggest that it may be misleading to think of individual neurons as tiny 'subjectivity pumps'. Indeed, if each neuron were to be seen as generating a quantum of conscious experience, one would then have to address problematic issues relating to the principles governing how quanta of consciousness add together (Ref. a). For instance, under what conditions would spatially disparate quanta add together? Would the quanta of neurons in more sophisticated circuits contribute more to the phenomenal sum than neurons in more primitive circuits?

Process theories avoid these questions because the activity of neural elements is bound together via computational considerations. Consciousness relates to global properties of the system, not to properties of its elements. One such global property is synchrony, where activity of a given neuron contributes to conscious content when it starts firing simultaneously with other neurons (Ref. b), perhaps those in some privileged processing structure (Ref. c). There are sound computational reasons for using synchrony to bind together different features of a single percept, thus solving the binding problem (Ref. d). The binding problem concerns the way in which the unity of conscious perception is achieved given that distinct aspects of the perceptual world are processed by functionally and anatomically distinct brain mechanisms. It can be illustrated by the following example: When one looks at a red square and a blue circle, color and shape must be bound together to disambiguate these stimuli from a blue square and a red circle. Yet the processing of different types of visual information, including color and shape, is performed by different systems in the brain. How does the right color information become associated with the right shape information? One answer is to have the neurons representing the color of a referent (e.g. red) generate firing that is synchronous with the firing representing the shape of that same referent (e.g. square).

Binding might also be the mechanism that accounts for the fact that phenomenal experience appears to be seamless and unified. In the above example, the experience of redness is not disconnected from the experience of squareness, nor is the experience of blueness disconnected from the experience of roundness. But this apparent unity raises many questions. Some theorists disagree that awareness is unified and so they don't require binding (Ref. e). Others argue that awareness can comprise disunified parts that can later be brought together into a unified experience by binding (Ref. f). Others argue that only the perceptual features that are bound can contribute to consciousness (Ref. b). Yet others argue that binding isn't necessary because a unified consciousness can emerge from the activity of the brain as a whole (Ref. g).

References

a Thomas, M.S.C. and Atkinson, A.P. (1999) Quantities of qualia. *Behav. Brain Sci.* 22, 169–170
b Tononi, G. and Edelman, G.M. (1998) Consciousness and complexity. *Science* 282, 1846–1851
c Bogen, J.E. (1998). Locating the subjectivity pump: the thalamic intralaminar nuclei. In *Toward a Science of Consciousness II: The 1996 Tucson Discussions and Debates* (Hameroff, S. *et al.*, eds), pp. 237–246, MIT Press
d Revonsuo, A. and Newman, J. (1999) Binding and consciousness. *Conscious. Cognit.* 8, 123–127
e O'Brien, G. and Opie, J. (1999) A connectionist theory of phenomenal experience. *Behav. Brain Sci.* 22, 127–196
f Zeki, S. and Bartels, A. (1999) Toward a theory of visual consciousness. *Conscious. Cognit.* 8, 225–259
g McCrone, J. (1999) *Going Inside: A Tour Round a Single Moment of Consciousness*, Faber and Faber

problem (see Box 4) and thereby be the neural correlate of visual awareness

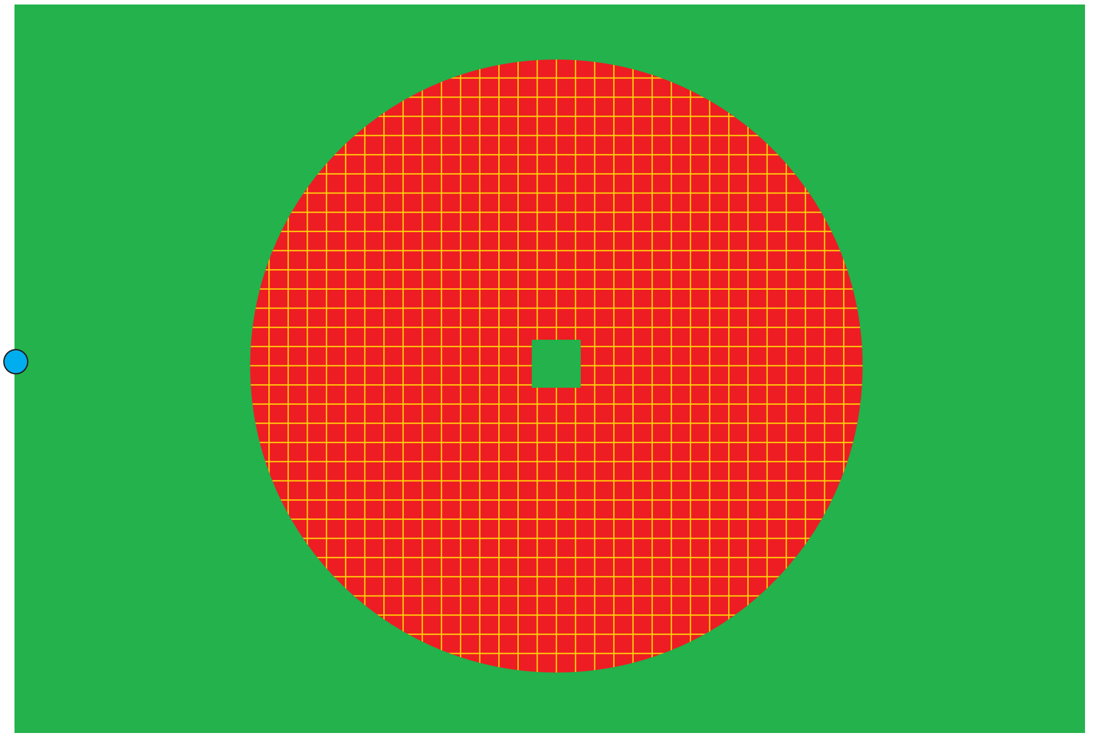
- Grossberg[48], for whom conscious processes are those that exhibit 'resonance', an integration of top-down and bottom-up information (see Box 5 for additional discussion of the extent to which consciousness involves high-level vs. low-level information).

Future directions

We have presented a distribution of theories across our conceptual landscape to illustrate the range of assumptions that have underpinned attempts to explain consciousness in computational terms. These theories are not equivalent: they differ not only in their emphases but also in the range of data they draw upon, in how encompassing they are and in how much evidential support they presently have. For example, the theories of Tononi and Edelman, and Crick and Koch, are drawn directly from neuroscientific data; the former is offered as a general theory of conscious experience while the latter is restricted to visual experience and is slightly more speculative. O'Brien and Opie, Greenfield, and Penrose and Hameroff offer general, but even more speculative, neuroscientific theories of consciousness. The theories of Schacter, Baars and Carruthers offer task-based functional analyses: Schacter's is established on the basis of neuropsychological dissociations, Baars' is drawn from both psychological and neuropsychological data and Carruthers' is based on philosophical concerns about what it is to be a thinking being. Converging neuroscientific evidence favors Baars' theory over the other two.

What can we conclude from the precise distribution of theories on our landscape? On the one hand, if weight of opinion is anything to go by, future explanations of consciousness will be process- rather than vehicle-based. On the other hand, there appears to be no clear trend as to whether these future theories will appeal to brain mechanisms dedicated to consciousness or not. Specialized

Box 5. Locating consciousness: does conscious experience correspond to the activity of 'low level' or 'high level' processing?



trends in Cognitive Sciences

Fig. I. (See text for explanation).

A visual illusion is shown in Fig. I. Hold the page ten inches from your face with the small blue circle central. Close your left eye and fixate on the blue circle. You should find that the green square at the center of the large red circle disappears (you may need to move the page towards or away from you slightly). When the green square has disappeared, ask yourself, 'Do I really experience the center of the red circle?' Not just 'Is the circle complete?' but 'Do I really see its center?'.

The process by which the green square is replaced by the grated red center is an example of 'filling in'. The square has fallen on the area of the retina known as the blindspot, where the optic nerve leaves the eye and there are no light receptors. The brain 'fills in' the missing information on the basis of the visual information around the blindspot. This information does not have to be a simple color, but may include complex patterns. For instance, filling in can occur when the information filled in is a dynamic twinkling pattern, like the static on an untuned TV screen (Ref. a).

Dennett has argued that there is no literal process of filling in (Ref. b). For the above diagram, when the square falls on the blindspot, the remaining visual input is sufficient for the observer to retrieve the high-level concept 'circle'. To retrieve the concept is to experience the circle, including its center. The brain does not re-instantiate the neural firings that would correspond to the low-level visual features associated with the center of the circle. In contrast, Churchland and Ramachandran have suggested that filling in of the blindspot occurs at a low level in the visual system (Ref. c) and have argued that such processes are fundamental in helping the visual system render conscious experience given noisy and incomplete input.

The difference between these theories concerns the level of abstraction at which processing becomes conscious or, in neural terms, those parts of the brain that generate consciousness (Refs d,e). Filling in is thus a useful empirical phenomenon in studying conscious visual awareness. It is still an open question whether conscious experience is associated with higher-order association areas or with perceptual systems themselves. While either can be cast as a process or a vehicle theory, the second position is more consistent with nonspecialized theories of consciousness. Crick and Koch (Refs f,g) have recently argued that early visual processing areas such as V1 do not contribute to consciousness. Their claim is based on evidence that the structure of perception does not map to the receptive field properties of V1 cells, and that V1 is not directly connected to frontal cortex where voluntary action is planned. Thus, a minimum level of abstraction may well be necessary before neural activity can generate conscious experience.

References

a Ramachandran, V.S. and Gregory, R.L. (1991) Perceptual filling in of artificially induced scotomas in human vision. *Nature* 350, 699–702

b Dennett, D.C. (1996) Seeing is believing – or is it? In *Vancouver Studies in Cognitive Science* (Vol. 5: *Perception*) (Akins, K.A., ed.), pp. 158–172, Oxford University Press

c Churchland, P.S. and Ramachandran, V.S. (1996) Filling in: why Dennett is wrong. In *Vancouver Studies in Cognitive Science* (Vol. 5: *Perception*) (Akins, K.A., ed.), pp. 132–157, Oxford University Press

d Akins, K.A. and Winger, S. (1996) Ships in the night: Churchland and Ramachandran on Dennett's theory of consciousness. In *Vancouver Studies in Cognitive Science* (Vol. 5: *Perception*) (Akins, K.A., ed.), pp. 173–197, Oxford University Press

e Pessoa, L. *et al.* (1998) Finding out about filling-in: a guide to perceptual completion for visual science and the philosophy of perception. *Behav. Brain Sci.* 21, 723–756

f Crick, F. and Koch, C. (1995) Are we aware of neural activity in primary visual cortex? *Nature* 375, 121–123

g Crick, F. and Koch, C. (1998) Consciousness and neuroscience. *Cereb. Cortex* 8, 97–107

mechanisms typically play a central role in explanations of A-consciousness, while non-specialized mechanisms are more often invoked to explain P-consciousness. This raises the interesting issue of whether the quadrants of our conceptual map are mutually exclusive across all types of consciousness (e.g. P-, A-, M- and S-). Explanations of distinct varieties of consciousness might fall in different quadrants. For example, if it turns out that felt awareness has nothing to do with computations *per se*, we might need a vehicle theory of P-consciousness and process theories of A-, M- and S-consciousness. Few theories address M-consciousness or S-consciousness beyond considerations of the relevance of these notions to the unified character of experience (Dennett's theory[11] is a notable exception). This is because in such models, self-knowledge and self-awareness are generally taken as just another kind of knowledge and another kind of awareness.

The models we have examined sometimes differ in the assumptions they make concerning two important issues: implementation and transitions.

First, while the distinction between vehicle and process theories depends on contrasting intrinsic properties of representations with the processes they are involved in, the very notion of a pure vehicle theory of consciousness sits uncomfortably in the information-processing framework. Indeed, pure vehicle theories must posit intrinsic properties of representations that are completely unrelated to processing considerations. It is unclear whether this is actually possible. Most current proposed properties (e.g. stability, above-threshold activation, global levels of firing) appear to be relevant only to the extent that the resulting representations can subsequently play a role in determining the future computational state of the system. What, then, distinguishes vehicle theories from process theories? Our view is that vehicle theories place relatively greater importance on implementation, that is, on the nature of the system performing the computations in generating consciousness. In contrast, process theories are more neutral concerning the nature of the substrate that implements the relevant computations, thereby encouraging the belief that any system implementing these computations will exhibit consciousness.

Second, the two poles of the specialized versus non-specialized dimension seem to require quite different accounts of transitions. By transitions, we mean the way in which consciousness emerges either in the evolution of organisms or in the development of individuals. One possible evolutionary scenario for non-specialized theories would be to assume that the emergence of consciousness depends on sheer structural complexity. Structural complexity is then taken to increase through evolution. One could likewise imagine similar non-specialized developmental accounts, whereby a growing organism does not achieve consciousness until the required structural complexity has developed. Specialized theories, on the other hand, necessarily have to account for the emergence of a dedicated mechanism for consciousness. In this respect, it is easy to imagine, for instance, how sharing information between domain-specific processing devices may be evolutionarily advantageous, if mechanisms providing such a facility afforded more flexible behavior in environments where selection pressures favored flexible behavior. Likewise, during development, going beyond mastery of a new skill to understand the structure of a problem and to use this information analogously in other domains is similarly advantageous[49]. Importantly, however, these arguments only apply to A-consciousness. It is much more difficult to develop similar scenarios to account for the emergence of P-consciousness (except under the view that pain isn't a real motivator unless it hurts!).

In conclusion, while future accounts of consciousness will undoubtedly be grounded in neuroscientific evidence, we believe that there is much to be gained from connecting the current search for the neural correlates of consciousness with a corresponding search for the 'compu-tational correlates of consciousness'. This nascent perspective, which we have surveyed in this review, offers the possibility of actively exploring the way in which different assumptions are consistent or inconsistent with each other, and allows evolutionary, developmental and information-processing considerations to be better integrated in the development of computational models of consciousness.

Outstanding questions

- Why and how does some physical matter produce subjective experience?
- Will an information-processing account of the brain produce a complete understanding of all, or only some, aspects of consciousness?
- What relation has animal consciousness to human consciousness?
- Is some binding process necessary for conscious perceptual experience? Is phenomenal experience necessarily unified?
- What is the correct way to describe phenomenal experience such that it can be directly associated with brain computations?
- What is the relation of natural selection to the computations/intrinsic representational properties that give rise to consciousness?
- Could any computational system that achieves the functionality of the human mind fail to be conscious?
- What is the appropriate level of description of neural activity such that it will correlate with the contents of consciousness? Will there necessarily be neural correlates of consciousness, or will the correlates be purely computational ones?

Acknowledgements

We are grateful to David Chalmers and the anonymous referees for their comments and advice. Thanks to Susan Greenfield for helpful discussions. Thanks also to the Department of Psychology, University of York, where A.P.A. was an Honorary Visiting Fellow during preparation of this article, and to King Alfred's College, for providing a travel grant. M.S.C.T. is supported by an MRC programme grant to Annette Karmiloff-Smith. A.C. is a research associate with the National Fund for Scientific Research (Belgium). His work was supported by a grant from the Université Libre de Bruxelles in support of IUAP program #P/4-19.

References

1 Chalmers, D.J. (1998) On the search for the neural correlate of consciousness. In *Toward a Science of Consciousness II: The 1996 Tucson Discussions and Debates* (Hameroff, S. *et al.*, eds), pp. 219–229, MIT Press

2 Chalmers, D.J. What is a neural correlate of consciousness? In *Neural Correlates of Consciousness: Conceptual and Empirical Questions* (Metzinger, T., ed.), MIT Press (in press)

3 Frith, C. *et al.* (1999) The neural correlates of conscious experience: an experimental framework. *Trends Cognit. Sci.* 3, 105–114

4 Block, N. (1995) On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–287
5 Young, A.W. and Block, N. (1996) Consciousness. In *Unsolved Mysteries of the Mind: Tutorial Essays in Cognition* (Bruce, V., ed.), pp. 149–179, Lawrence Erlbaum
6 Nagel, T. (1974) What is it like to be a bat? *Philos. Rev.* 83, 435–450
7 Chalmers, D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press
8 Levine, J. (1983) Materialism and qualia: the explanatory gap. *Pacific Philos. Q.* 64, 354–361
9 Levine, J. (1993) On leaving out what it's like. In *Consciousness: Psychological and Philosophical Essays* (Davies, M. and Humphreys, G.W., eds), pp. 121–136, Basil Blackwell
10 Dennett, D.C. (1988) Quining qualia. In *Consciousness in Contemporary Science* (Marcel, A.J. and Bisiach, E., eds), pp. 42–77, Clarendon Press
11 Dennett, D.C. (1991) *Consciousness Explained*, Little, Brown
12 Kurthen, M. *et al.* (1998) Will there be a neuroscientific theory of consciousness? *Trends Cognit. Sci.* 2, 229–234
13 O'Brien, G. and Opie, J. (1999) A connectionist theory of phenomenal experience. *Behav. Brain Sci.* 22, 127–196
14 Mangan, B. (1993) Dennett, consciousness and the sorrows of functionalism. *Conscious. Cognit.* 2, 1–17
15 Mathis, W.D. and Mozer, M.C. (1996) Conscious and unconscious perception: a computational theory. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, pp. 324–328, Lawrence Erlbaum
16 Kinsbourne, M. (1997) What qualifies a representation for a role in consciousness? In *Scientific Approaches to Consciousness* (Cohen, J.D. and Schooler, J.W., eds), pp. 335–355, Lawrence Erlbaum
17 Tononi, G. and Edelman, G.M. (1998) Consciousness and complexity. *Science* 282, 1846–1851
18 Van Gelder, T. (1998) The dynamical hypothesis in cognitive science. *Behav. Brain Sci.* 21, 615–665
19 Bechtel, W. and Richardson, R.C. (1993) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton University Press
20 Cummins, R. (1983) *The Nature of Psychological Explanation*, MIT Press
21 Ellis, A.W. and Young, A.W. (1996) *Human Cognitive Neuropsychology: A Textbook with Readings*, Psychology Press
22 Young, A.W. (1998) *Face and Mind*, Oxford University Press
23 Hirschfeld, L.A. and Gelman, S.A. (eds) (1994) *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge University Press
24 Barkow, J.H. *et al.*, eds (1992) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, Oxford University Press
25 Atkinson, R.C. and Shiffrin, R.M. (1971) The control of short-term memory. *Sci. Am.* 224, 82–90
26 Norman, D.A. and Shallice, T. (1986) Attention to action: willed and automatic control of behavior. In *Consciousness and Self Regulation: Advances in Research* (Vol. 4) (Davidson, R.J. *et al.*, eds), pp. 1–18, Plenum Press
27 Schacter, D.L. (1989) On the relations between memory and consciousness: dissociable interactions and conscious experience. In *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving* (Roedigger, H.L., III and Craik, F.I.M., eds), pp. 355–389, Lawrence Erlbaum
28 Baars, B.J. (1988) *A Cognitive Theory of Consciousness*, Cambridge University Press
29 Atkinson, A.P. and Thomas, M.S.C. (1999) What makes us conscious? *J. Intell. Syst.* 9, 307–354
30 Duncan, J. (1995) Attention, intelligence, and the frontal lobes. In *The Cognitive Neurosciences* (Gazzaniga, M.S., ed.), pp. 721–733, MIT Press
31 Rabbitt, P., ed. (1997) *Methodology of Frontal and Executive Function*, Psychology Press
32 Tranel, D. *et al.* (1994) Development of the concept of 'executive function' and its relationship to the frontal lobes. In *Handbook of Neuropsychology* (Vol. 9) (Boller, F. and Spinnler, H., eds), pp. 125–148, Elsevier
33 Baars, B.J. (1995) Evidence that phenomenal consciousness is the same as access consciousness. *Behav. Brain Sci.* 18, 249
34 Baars, B.J. and Newman, J. (1994) A neurobiological interpretation of global workspace theory. In *Consciousness in Philosophy and Cognitive Neuroscience* (Revonsuo, A. and Kamppinen, M., eds), pp. 211–226, Lawrence Erlbaum
35 Newman, J. *et al.* (1997) A neural Global Workspace model for conscious attention. *Neural Netw.* 10, 1195–1206
36 Carruthers, P. (1996) *Language, Thought and Consciousness: An Essay in Philosophical Psychology*, Cambridge University Press
37 Rolls, E.T. (1998) *The Brain and Emotion*, Oxford University Press
38 Rosenthal, D.M. (1986) Two concepts of consciousness. *Philos. Stud.* 49, 329–359
39 Rosenthal, D.M. (1993) Thinking that one thinks. In *Consciousness: Psychological and Philosophical Essays* (Davies, M. and Humphreys, G.W., eds), pp. 197–223, Basil Blackwell
40 Thomas, M.S.C. and Atkinson, A.P. (1999) Quantities of qualia. *Behav. Brain Sci.* 22, 169–170
41 Van Gulick, R. (1999) Vehicles, processes, and neo-classical revival. *Behav. Brain Sci.* 22, 170–171
42 Greenfield, S. (1995) *Journey to the Centres of the Mind*, W.H. Freeman
43 Penrose, R. and Hameroff, S. (1995) What gaps? *J. Conscious. Stud.* 1, 17–24
44 Churchland, P.S. (1998) Brainshy: nonneural theories of conscious experience. In *Toward a Science of Consciousness II: The 1996 Tucson Discussions and Debates* (Hameroff, S. *et al.*, eds), pp. 109–126, MIT Press
45 Edelman, G.M. (1989) *The Remembered Present: A Biological Theory of Consciousness*, Basic Books
46 Crick, F. and Koch, C. (1990) Towards a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275
47 Crick, F. and Koch, C. (1992) The problem of consciousness. *Sci. Am.* 267, 110–117
48 Grossberg, S. (1999) The link between brain learning, attention, and consciousness. *Conscious. Cognit.* 8, 1–44
49 Karmiloff-Smith, A. (1992) *Beyond Modularity: A Developmental Perspective on Cognitive Science*, MIT Press