Consciousness: converging insights from connectionist modeling and neuroscience

Tiago V. Maia[1] and Axel Cleeremans[2]

[1]Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
[2]Cognitive Science Research Unit, Université Libre de Bruxelles CP 191, Avenue F.-D. Roosevelt, 50, 1050 Bruxelles, Belgium

Over the past decade, many findings in cognitive neuroscience have resulted in the view that selective attention, working memory and cognitive control involve competition between widely distributed representations. This competition is biased by top-down projections (notably from prefrontal cortex), which can selectively enhance some representations over others. This view has now been implemented in several connectionist models. In this review, we emphasize the relevance of these models to understanding consciousness. Interestingly, the models we review have striking similarities to others directly aimed at implementing 'global workspace theory'. All of these models embody a fundamental principle that has been used in many connectionist models over the past twenty years: global constraint satisfaction.

Introduction

We believe that a view that allows the integration of selective attention, working memory, cognitive control and consciousness is within sight. This view is based on the notion of biased competition: competition between representations that are widely distributed in the brain, with top-down influences, most notably from prefrontal cortex [1–3]. This mechanism is now implemented in several connectionist models that collectively address a wide range of findings [4–10]. Many of these models have not been directly targeted at explaining consciousness, but they have important implications for our understanding of conscious information processing. In this article, we review these developments, highlight key computational principles embodied in these models, and explore their implications for consciousness. We also emphasize that these models can be understood in terms of a fundamental principle that can be traced back to early connectionist efforts: global constraint satisfaction [11].

It is useful to distinguish between *states* of consciousness (e.g. being awake, asleep, in a coma, etc.) and the *contents* of consciousness (e.g. being conscious of the scene one is looking at) [12] (although it should be noted that these are not independent [13]). The present article is

about the contents of consciousness: we will not address what might be called the 'enabling factors' for consciousness (e.g. appropriate neuromodulation from the brainstem, etc.).

Box 1 highlights some of the computational mechanisms that we hypothesize are associated with the contents of consciousness. These are developed in the remainder of this review.

Consciousness and active representations

When talking about biological or artificial neural networks, it is important to distinguish between two types of representation. First, there is long-term knowledge that is embedded (or latent) in the weights of the connections between units. This knowledge can drive behavior

---

Box 1. Mechanisms hypothesized to be associated with consciousness

(1) Active representation
Active neuronal firing is necessary (but probably not sufficient) for consciousness.

(2) Global competition biased by top-down modulation
Consciousness results from global competition between representations. At any moment, the winning neuronal coalition determines both conscious phenomenal experience and global accessibility. Active representations maintained by PFC are important sources of biases for this competition.

(3) Global constraint satisfaction
Global competition implements global constraint satisfaction. Thus, conscious experience can be seen as the result of a large-scale application of the brain's knowledge to the current situation.

(4) Reentrant processing
Recurrent connections are essential to implement global constraint satisfaction. They allow more global interpretations in higher-level areas to influence processing in lower-level areas (which tend to work more like localized feature detectors).

(5) Meta-representation
Higher aspects of human consciousness and cognition, such as the ability to think about one's thoughts, may depend on the creation of representations that are then fed back to the same constraint satisfaction network as input [11,71]. More generally, the creation of representations that are then available for re-processing by the same network implements a cycle that could be the basis of the 'stream of thought'.

Corresponding author: Cleeremans, A. (axcleer@ulb.ac.be).

indirectly by eliciting specific firing patterns over ensembles of units, but does not seem to be available for *direct* inspection by, or transmission to, another system. Although some have criticized this aspect of connectionist models, we believe that it helps to explain implicit cognition [14]. For example, most humans are proficient users of their native language but are hardly able to enumerate any grammatical rules without explicit instruction; we suggest that this is because grammar is encoded in the connections.

In addition to knowledge embedded in the connections, one also finds typically more transient, active representations, in the form of firing patterns. If, as we hypothesize, the knowledge in the connections is not directly accessible, conscious representations must depend on these active representations [12]. This leads us to hypothesize that only the *outputs* of computations in the brain (possibly including intermediate results) can potentially be conscious, whereas the mechanisms of the computations themselves remain impenetrable (see also [15]). It is important to emphasize, though, that we are not claiming that the active firing of neurons is *sufficient* for consciousness, or that the outputs of every computation are conscious. Rather, we propose that these may be *necessary* conditions for a representation to become conscious (see, e.g. [12,16] for similar suggestions).

Some results suggest that neural firing may need to be sufficiently strong and sustained through time (i.e. stable [17,18]) to support a conscious experience. For example, cortical microstimulation studies in sensory areas in humans show that for conscious perception, stimulation of sufficient duration and intensity is required [19]. Similarly, brief or weak presentations of stimuli might result in subliminal perception, whereas longer or stronger presentations would result in conscious perception [20]. There is also evidence from neuroimaging that consciousness is associated with increased firing [21]. However, few would argue that strong and sustained neural firing is *sufficient* for a conscious experience (but see [22]). What else is needed?
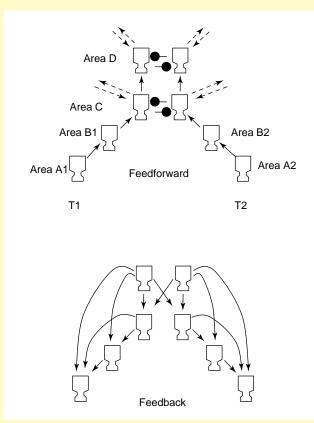
Global workspace theory

Baars has suggested that consciousness depends on access to a 'global workspace' [23] and Dehaene and collaborators have developed several neurocomputational models that implement this theory (e.g. [24,25]). A crucial feature of these models is that they work by biased competition (Box 2). The basic assumption is that the winning coalition of neurons determines conscious experience at a given moment. Many theorists (e.g. Crick and Koch [26], Edelman [27]) have made similar proposals, and this is also our working hypothesis. The main difference between our perspective and that of Dehaene, Baars, and their collaborators is that they take the brain to consist of specialized, modular processors and a global workspace that connects these processors, whereas we believe that computation is more distributed and interactive at a global scale. In particular, the existence of massive recurrent connections at all levels of the cortex makes the existence of strongly encapsulated modules, as suggested by Dehaene and colleagues (e.g. [24]), unlikely.
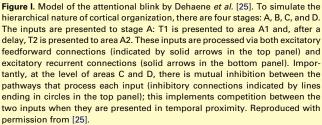

Box 2. Neurocomputational models of global workspace theory

Dehaene and collaborators have implemented several models of global workspace theory. As an example, we examine a model that they have used to simulate the attentional blink [25]. In this paradigm, participants are presented with two successive stimuli; if the interval between the stimuli is short, participants' ability to report the second stimulus decreases (as if their attention 'blinks').

Figure I shows the architecture of the model. The model exhibits the attentional blink and can account for several neurophysiological findings [25]. The basic idea is that when stimulus T1 is presented, the network creates a global state in which T1 is represented at all levels of the hierarchy. Because of the recurrent connections, this state is self-sustained for a short period. If T2 is presented shortly after T1, it faces the competition of the lingering representation of T1 in areas C and D and cannot be effectively processed.

The key principles embodied in the model are lateral inhibition, which implements competition in all the models we review, and recurrent connectivity. Interestingly, in this model the top-down influences are just the result of lingering activation from the previous stimulus. Other models that we review include top-down biases that represent the focus of attention (e.g. [7–10]; Box 4), items in working memory (e.g. [3,5]; Box 3) or task demands (e.g. [8,31]). This model seems to be missing an important component included in these other models: sustained top-down biases from PFC. Indeed, Dehaene *et al.* note that this model would need to include top-down attentional biases to account for a broader range of empirical results [25], and they implement such biases in other work [24].

Figure I. Model of the attentional blink by Dehaene *et al.* [25]. To simulate the hierarchical nature of cortical organization, there are four stages: A, B, C, and D. The inputs are presented to stage A: T1 is presented to area A1 and, after a delay, T2 is presented to area A2. These inputs are processed via both excitatory feedforward connections (indicated by solid arrows in the top panel) and excitatory recurrent connections (solid arrows in the bottom panel). Importantly, at the level of areas C and D, there is mutual inhibition between the pathways that process each input (inhibitory connections indicated by lines ending in circles in the top panel); this implements competition between the two inputs when they are presented in temporal proximity. Reproduced with permission from [25].

In any case, this may be simply a matter of emphasis, as Dehaene *et al.* suggest that 'global workspace neurons' are widely distributed [24].

The idea that the winning coalition in a large-scale competition determines the contents of consciousness can explain the association between consciousness and strength and stability of neuronal firing: strong and sustained firing makes it more likely that the corresponding representation will be part of the winning neuronal coalition, and conversely, neurons that are part of a winning coalition receive excitation from other neurons in the coalition, which could result in an increased firing rate. These observations are related to Varela and Thompson's notion of local-to-global and global-to-local causality [28], and explain, for example, the fact that information that does not enter consciousness tends to decay quickly [29].

The dynamics of the competition between neuronal coalitions gives rise to an interesting prediction, which has been confirmed experimentally: that access to consciousness, although based on graded processing, tends to be 'all-or-none' [13,25] (but see [14]). This follows from the nonlinear dynamics of lateral inhibition [30]. Ongoing spontaneous activity in the thalamocortical system may also be important in determining the winning coalition [13].

Next, we turn to models that have not been directly targeted at consciousness, but that address closely related issues: attention, working memory and cognitive control. We will see that these models also work by biased competition and that attention, working memory, cognitive control and consciousness can be understood in terms of the dynamics of this competition.

An integrated view of attention, working memory, cognitive control and consciousness
Attention, working memory and cognitive control
The idea of large-scale competition biased by top-down projections from prefrontal cortex (PFC) has recently been proposed as the mechanism underlying attention [1], working memory [3], and cognitive control [2,6,31]. Braver, Cohen, O'Reilly and colleagues have implemented these ideas in several connectionist models of working memory and cognitive control (see Box 3). There have also been several connectionist implementations of biased competition models of selective attention (see Box 4).

The PFC plays a crucial role in all of these models. It appears to be specialized for maintaining active representations and, working together with the basal ganglia and/or the dopaminergic system, has the ability to switch rapidly between representations when necessary [2–5,32,33]. This is crucial for the ability to maintain and update representations in working memory, change the representation of task demands in cognitive control, or modify the focus of attention, flexibly and quickly. O'Reilly and colleagues have proposed specific mechanisms that

---

Box 3. Biased competition models of working memory and cognitive control

Braver, Cohen, O'Reilly and colleagues have developed several models of working memory and cognitive control that are based on biased competition (e.g. [3–6,8,31,32,72]). As an example, we examine a model they have used to address findings with the AX–CPT task (Figure I). In this task, participants see a stream of letters and have to press one button (in the case of the model in Figure I, the left button) when they see an X following an A, and another button (right) for every other letter. This is a simple working memory task because participants have to keep the last stimulus in working memory, so that when an X appears they know whether an A preceded it.

In the model, posterior perceptual and motor cortex (PMC) maps the stimuli to the responses. However, this mapping is influenced by top-down projections from PFC, which remembers whether the last stimulus was an A. Therefore, the response is influenced both by the environmental stimulus and the contents of working memory (stored in PFC). On the other hand, presentation of a stimulus elicits not only a response but also an update to working memory. (In this simple model, every new stimulus enters working memory, thereby erasing the previous stimulus. In more complex models, the system learns when it should update working memory and when it should maintain its current representation despite incoming distractors).

Two features of the model are of special interest here. First, there is lateral inhibition in each layer. As in the other models we review, this implements competition. Second, the PFC layer is capable of biasing competition 'downstream' (via its projections to PMC) and maintaining activation during a delay (through recurrent loops not shown). The model is an excellent example of biased competition.
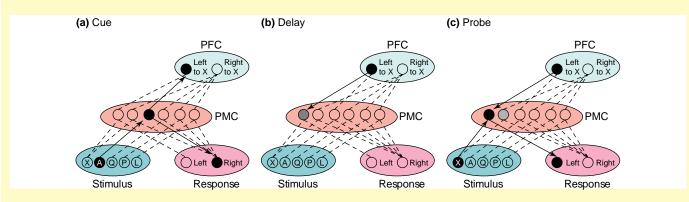

Figure I. Model of the AX-CPT task by Cohen, Braver and O'Reilly. PFC, prefrontal cortex. PMC, posterior perceptual and motor cortex. (a) When an A is presented, it gets mapped onto a 'Right' response, but it also activates the PFC unit ('Left to X') that indicates that A was just presented. (b) During the delay period, activity in PFC is sustained, and this provides top-down activation to the unit that maps X to the 'Left' response. (c) When X is presented after the delay, it activates two units in PMC, but the one that indicates the 'Left' response receives further support from the top-down bias from PFC and hence wins the competition; the model therefore makes the correct response ('Left'). Adapted with permission from [3].

Box 4. Biased competition models of selective attention

Several connectionist models of selective attention are based on the principle of biased competition (e.g. [7–10]). As an example, we briefly examine a model that Corchs and Deco have used to simulate the neurophysiological results of two experiments on selective visual attention ([9]; see Figure I). There are two main pathways in the model, corresponding to the ventral and dorsal visual pathways in the brain. The ventral pathway goes from V1 to the extrastriate cortex (areas V2-V4) and then to inferior temporal cortex (IT), and in the model implements object recognition. The dorsal pathway goes from V1 and extrastriate cortex to posterior parietal cortex (PP), and in the model codes the location of objects. (There has been much discussion about the functions of the dorsal pathway, and some authors disagree with its characterization as the 'where' pathway. However, this need not concern us here: our purpose is to use this

model to illustrate how one can understand selective attention in terms of biased competition, not to shed light on the detailed organization of the visual system).

All layers in the model have lateral inhibition (implemented by the inhibitory pools). As in the other models we review, this creates competition. There are also extensive feedback projections from higher- to lower-level areas, as well as top-down projections from PFC that implement selective attention by biasing the competition in favor of a particular location (through the projection from dorsal area 46 to PP) or object (through the projection from ventral area 46 to IT). Note that although attention enhances certain representations, it cannot guarantee that those representations win the global competition – salient stimuli might still win the competition even if they are outside the focus of attention.
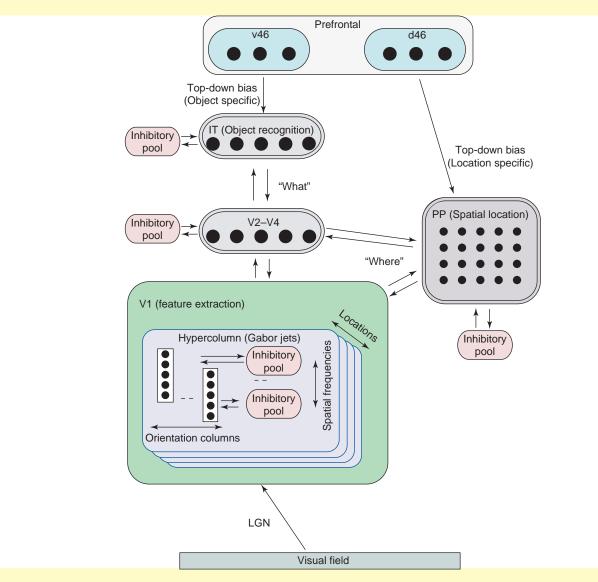

Figure I. Model of selective visual attention by Corchs and Deco [9]. LGN, lateral geniculate nucleus. IT, inferior temporal cortex. PP, posterior parietal cortex. v46, ventral area 46 (dorsolateral prefrontal cortex). d46, dorsal area 46 (dorsolateral prefrontal cortex). See text for details. Adapted with permission from [9].

support this flexibility and have implemented those mechanisms in computational models (e.g. [4]). The basic idea is that recurrent excitation (or several other mechanisms [4]) maintains activity in PFC neurons; a 'gating' mechanism implemented by the basal ganglia (or the dopaminergic system [6,34]) bars access to the PFC

when it is necessary to maintain information in PFC without interference, and quickly allows access to PFC when it is necessary to modify or replace that information [4].

The general idea, then, is that representations that are actively maintained in PFC help to bias the competition between representations elsewhere in the brain, in ways

that are appropriate for the current goals and task. It should be noted that, in a sense, competition in a given area is 'biased' by all of the inputs to that area – thus, competition in V4 could be said to be 'biased' by feedforward projections from V2 to V4 or feedback projections from IT to V4. What seems to be special about PFC is that it has the ability to *sustain* representations in the absence of input and can therefore implement biases that are sustained over time, even in the face of distractors. Some lower areas, such as IT, seem to have some ability to maintain a representation active during a delay, but that representation is overridden as soon as a new input is processed [2]; these areas cannot therefore be the source of the sustained biases necessary for attention, working memory and cognitive control.

The fact that biased competition seems to underlie attention, working memory and cognitive control suggests an integration of these concepts. This is particularly evident if one notices the similarities in the operation of the models that have been proposed to address findings in these domains (see Boxes 3 and 4). However, it has also not escaped the attention of neuroscientists; for example, Courtney has recently proposed that attention and cognitive control are 'emergent properties of information representation in working memory' ([35], p. 501), noting that 'attention and cognitive control arise as a consequence of WM [working memory], but are not separate entities that are specifically implemented by the brain' ([35], p. 503). In other words, representations that are actively maintained in working memory bias the global competition in ways 'that we call attention and cognitive control' ([35], p. 503). These ideas are compatible with the proposals of several theorists (e.g. [1–3]).

Consciousness

The relevance of these models for consciousness is suggested both by their similarity to models that have been explicitly aimed at implementing global workspace theory (see Box 2) and by the close relation between consciousness and attention, working memory and cognitive control. Consciousness is widely assumed to be intimately tied to attention [29], in that unattended stimuli often fail to enter consciousness [36] (but see [37]). It is also closely related to cognitive control; in fact, the notion of 'controlled processing' has been associated with effortful and voluntary processing 'under control of […] the subject' ([38], p. 2) since it was proposed [38,39]. Finally, consciousness is also closely related to working memory (e.g. [40,41]).

The crucial role that the PFC plays in all of these processes and in biasing activation elsewhere in the brain [2] suggests that it probably plays an important role in selecting and maintaining the contents of consciousness (see, e.g. [3,29]). Additional evidence confirms the likely importance of the PFC for consciousness (e.g. [16,42]). For example, conscious stimuli typically activate frontal regions whereas non-conscious stimuli do not; in unconscious states such as coma or general anesthesia sensory stimulation activates sensory cortex but not frontal regions; and frontal regions show marked metabolic decrements in unconscious states [42]. Furthermore,

deficits in PFC lead to difficulties selecting and sustaining task-relevant representations, as would be expected. For example, lesions to the PFC often result in high distractibility [43,44] and the PFC is prominently implicated in attention-deficit hyperactivity disorder [43].

Theoretical and methodological implications

In short, our suggestion is that attention, working memory, cognitive control and consciousness are *not* distinct functions implemented by separate brain systems. Attempting to find separate neural correlates for each may therefore be the wrong approach. Instead, we suggest that they should be understood in terms of the dynamics of global competition, with biasing from PFC. Methodologically, this points to the importance of global measures of brain function that can quantify the state of the global competition; Tononi and Edelman's measures of 'integration' ('functional clustering') and 'differentiation' ('neural complexity') [45] are excellent examples of the new kinds of tools that may be needed for this endeavor.

Note that the framework we suggest naturally answers questions that have plagued the literature on consciousness. For example, there has been much discussion about the relation between attention and consciousness: whether attention is necessary for consciousness, whether attention precedes consciousness or vice-versa, etc. (e.g. [37]). Under the perspective articulated here, these are simply the wrong questions. A stimulus enters consciousness if its representation is part of the winning coalition. This is modulated, but by no means fully determined, by top-down biases from PFC. One may choose to call these biases 'attention'; the key issue, however, is that a thorough understanding of the processes involved requires understanding the overall dynamics of the global competition.

Reentrant connections and consciousness

We have seen that the PFC probably influences the contents of consciousness via top-down projections. More broadly, the importance of feedback (or reentrant) connections for conscious experience has been emphasized by several authors [25,27,37], and there is evidence from neuroscience to support these views (see, e.g. [37] for review). For example, transcranial magnetic stimulation of visual area MT induces motion perception, but not if activity in V1 is disrupted some time after the stimulation of MT [46].

Consciousness and global constraint satisfaction

The idea that recurrent interactions at a nearly global scale are important for consciousness [25,27,37,47,48] makes perfect sense in light of the idea that neural networks implement global constraint satisfaction – an idea that goes back to the early days of connectionism [49]. Many early connectionist models used global constraint satisfaction to address findings in several domains (e.g. [11,50,51]). The idea is that a network with recurrent connections arrives at an interpretation of a given input by settling into a stable state [49]. This state is a function not only of the network's input but also of the knowledge embedded in the network's connections (this is why we

talk about an 'interpretation' of the input). We suggest that conscious experience reflects stable states that correspond to interpretations that the brain makes of its current inputs (see also [11,24,25]). (These inputs can include representations that were generated internally in a previous time step, as in inner speech. Unfortunately we do not have space to expand on that idea here, but see [11].) We suppose that, in many cases, the relevant constraints are spread throughout much of the brain. This explains the need for massive global interactions to reach a stable state that supports a given conscious experience.

Mathis and Mozer have explored the relation between consciousness and constraint satisfaction and implemented these ideas in connectionist models [52]. Their approach differs from ours in that, like Dehaene and collaborators [24], they believe that there are multiple interconnected modules; they further assume that constraint satisfaction occurs *within* each module. We, on the other hand, emphasize large-scale (*global*) constraint satisfaction. This seems more consistent with the neuroanatomical evidence of extensive recurrent connectivity and also with behavioral findings, such as the McGurk effect [53], in which perception in one modality is influenced by stimuli presented in another modality. These ideas of the relation between consciousness and global constraint satisfaction find an interesting echo in Perruchet and Vinter's proposal that conscious experience is self-organizing so as to best reflect the structure of the environment [54]. The notion of conscious experience as a sequence of large-scale stable states [11,26,27,29] also explains how a massively parallel brain can produce a seemingly serial 'stream of consciousness' [55].

Synchronization, binding and global competition
There has been much debate about the role of synchronized neuronal firing to solve the so-called 'binding problem' [56] and determine the contents of consciousness (e.g. [26,57–59]). This literature is too extensive to cover in detail here, but we would like to emphasize that the 'synchrony hypothesis' is fundamentally compatible with our proposals, as competition need not be implemented exclusively in terms of firing rate: synchronization may also play an important role.

Consistent with this idea, the main proponents of the synchrony hypothesis now believe that synchronized firing is not a sufficient condition for consciousness [26,58] and see it instead as important in influencing which coalition wins the global competition. For example, Crick and Koch recently stated that 'we no longer think that synchronized firing […] is a sufficient condition for the NCC. One likely purpose of synchronized firing is to assist a nascent coalition in its competition with other (nascent) coalitions' ([26], p. 123); Engel and Singer [58] make similar points. Synchronized firing thus seems intimately related to global competition.

Several neurocomputational models of the role of synchronization in perception and attention (e.g. [60–62]) and even reasoning [63] have been proposed. In fact, von der Malsburg used temporal synchrony in his models [64,65] to solve what later became known as the binding problem several years before the consciousness community started to focus on this problem. An important area for future research is to understand better the relation between these models and models that implement global competition based on firing rate. As Fries *et al.* noted, 'synchronization is likely to be translated into firing rate changes at later processing stages' ([66], p. 3739; see also [58]); thus, synchronization and firing rate are likely to be intimately intertwined. Connectionist modeling has the potential to shed light on this relation (see, e.g. [67]), and possibly start to address the independent or joint contributions of firing rate and synchronization to consciousness (see also Box 5).

Access versus phenomenal consciousness
A decade ago, Block proposed a distinction between 'access' and 'phenomenal' consciousness [68]. In relation to the present framework, our working hypothesis is that the winning coalition determines both global accessibility *and* phenomenal experience, as a large-scale stable state seems an appropriate candidate for global access and would also explain the integrated character of phenomenal experience [45]. Block has recently suggested that losing coalitions might also determine the contents of phenomenal experience, arguing specifically that, for example, losing coalitions in recurrent MT/V1 loops might be sufficient for phenomenal consciousness [69]. However, in a recent review, Tong concluded that 'activity throughout the ventral extrastriate pathway […] does not seem to be sufficient for awareness' [70]. Furthermore, the notion that losing neuronal coalitions might contribute to phenomenal experience seems at odds with the integrated and unitary character of such experience: people do not normally seem to have disjointed perceptual experiences, with the intrusion of movements, lines, and so on, that correspond to the activity of losing coalitions. Notably, in binocular rivalry or ambiguous visual stimuli (e.g. the Necker cube) people usually report that they perceive one

> **Box 5. Outstanding questions and directions for future research**
>
> • What is the relation between competition based on synchrony and competition based on firing rates? When is one or the other used, and how are they interrelated?
> • What about qualia (the 'hard problem' of consciousness [73])? It seems that it would be possible to have a system without qualia that would have the properties we outlined. What is missing from the explanation?
> • If consciousness is a sequence of 'snapshots', or stable states, what accounts for the integrated, continuous flux of phenomenal experience? In his recent book, Lloyd explores the problem of temporality in consciousness in great detail [74]. He relates ideas regarding temporality in phenomenal experience (that go back to Husserl) to implementations using recurrent neural networks and even fMRI findings. An exciting area of future research will be to integrate those ideas with the framework presented here.
> • An important area for future research is to attempt to build an integrated neurocomputational architecture that can address findings in attention, working memory, cognitive control and consciousness with minimal individual customization. A possible starting point is the Leabra framework of O'Reilly and collaborators [75], which has been used to implement several of the models of working memory and cognitive control discussed in this article.

interpretation or the other – not a mixture of the two. This is exactly what would be predicted by a global constraint satisfaction account [11]. (Of course it could be argued that people do constantly experience disjoint representations but do not report them because such representations are not available to access consciousness. Based on our own phenomenal experience, we find it more parsimonious to assume that this is not the case until concrete evidence suggests otherwise.)

Conclusions

In summary, recent and current work in connectionist modeling and neuroscience is converging to provide an integrated view of attention, working memory, cognitive control and consciousness based on a single mechanism: global competition between representations, with top-down biases from PFC. This fosters an integrated understanding of these concepts in terms of the mechanisms and dynamics of global competition, rather than as reified processes with distinct neural instantiations.

References

1 Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222

2 Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202

3 O'Reilly, R.C. *et al*. (1999) A biologically based computational model of working memory. In *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (Miyake, A. and Shah, P., eds), pp. 375–411, Cambridge University Press

4 Frank, M.J. *et al*. (2001) Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cogn. Affect. Behav. Neurosci.* 1, 137–160

5 O'Reilly, R.C. (2003) *Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia*. Technical Report (03-03), Institute of Cognitive Science, University of Colorado at Boulder

6 Braver, T.S. and Cohen, J.D. (2000) On the control of control: The role of dopamine in regulating prefrontal function and working memory. In *Control of Cognitive Processes: Attention and Performance XVIII* (Monsell, S. and Driver, J., eds), pp. 713–737, MIT Press

7 Deco, G. and Rolls, E.T. (2005) Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons. *J. Neurophysiol.* (Articles in press: No. 10952004)

8 Cohen, J.D. *et al*. (1990) On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol. Rev.* 97, 332–361

9 Corchs, S. and Deco, G. (2002) Large-scale neural model for visual attention: Integration of experimental single-cell and fMRI data. *Cereb. Cortex* 12, 339–348

10 Spratling, M.W. and Johnson, M.H. (2004) A feedback model of visual attention. *J. Cogn. Neurosci.* 16, 219–237

11 Rumelhart, D.E. *et al*. (1986) Schemata and sequential thought processes in PDP models. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2) (McClelland, J.L. *et al*., eds), pp. 7–57, MIT Press

12 Dehaene, S. and Changeux, J.P. (2004) Neural mechanisms for access to consciousness. In *The Cognitive Neurosciences* (Gazzaniga, M.S., ed.), pp. 1145–1158, MIT Press

13 Dehaene, S. and Changeux, J.P. (2005) Ongoing spontaneous activity controls access to consciousness: A neuronal model for inattentional blindness. *PLoS Biol.* 3, e141

14 Cleeremans, A. Computational correlates of consciousness. *Prog. Brain Res.* (in press)

15 Velmans, M. (1991) Is human information processing conscious? *Behav. Brain Sci.* 14, 651–726

16 Koch, C. (2004) *The Quest for Consciousness: A Neurobiological Approach*, Roberts & Co.

17 Cleeremans, A. *Being Virtual*, Oxford University Press (in press)

18 Cleeremans, A. and Jimenez, L. (2002) Implicit learning and consciousness: A graded, dynamic perspective. In *Implicit Learning and Consciousness: An Empirical, Computational and Philosophical Consensus in the Making?* (French, R.M. and Cleeremans, A., eds), pp. 1–40, Psychology Press

19 Libet, B. *et al*. (1964) Production of threshold levels of conscious sensation by electrical stimulation of human somatosensory cortex. *J. Neurophysiol.* 27, 546–578

20 Merikle, P.M. *et al*. (2001) Perception without awareness: perspectives from cognitive psychology. *Cognition* 79, 115–134

21 Moutoussis, K. and Zeki, S. (2002) The relationship between cortical activation and perception investigated with invisible stimuli. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9527–9532

22 O'Brien, G. and Opie, J. (1999) A connectionist theory of phenomenal experience. *Behav. Brain Sci.* 22, 127–148

23 Baars, B.J. (1988) *A Cognitive Theory of Consciousness*, Cambridge University Press

24 Dehaene, S. *et al*. (1998) A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14529–14534

25 Dehaene, S. *et al*. (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8520–8525

26 Crick, F. and Koch, C. (2003) A framework for consciousness. *Nat. Neurosci.* 6, 119–126

27 Edelman, G.M. (2003) Naturalizing consciousness: a theoretical framework. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5520–5524

28 Varela, F.J. and Thompson, E. (2003) Neural synchrony and the unity of mind: a neurophenomenological perspective. In *The Unity of Consciousness: Binding, Integration, and Dissociation* (Cleeremans, A., ed.), pp. 266–287, Oxford University Press

29 Dehaene, S. and Naccache, L. (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79, 1–37

30 Wilson, H.R. (1999) *Spikes, Decisions, and Actions: The Dynamical Foundations of Neuroscience*, Oxford University Press

31 Cohen, J.D. *et al*. (2004) A systems-level perspective on attention and cognitive control: Guided activation, adaptive gating, conflict monitoring, and exploitation versus exploration. In *Cognitive Neuroscience of Attention* (Posner, M.I., ed.), pp. 71–90, Guilford Press

32 Atallah, H.E. *et al*. (2004) Hippocampus, cortex, and basal ganglia: insights from computational models of complementary learning systems. *Neurobiol. Learn. Mem.* 82, 253–267

33 Goldman-Rakic, P.S. (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In *Handbook of Physiology: The Nervous System* (Vol. 5) (Plum, F., ed.), pp. 373–417, American Physiological Society

34 Montague, P.R. *et al*. (2004) Computational roles for dopamine in behavioural control. *Nature* 431, 760–767

35 Courtney, S.M. (2004) Attention and cognitive control as emergent properties of information representation in working memory. *Cogn. Affect. Behav. Neurosci.* 4, 501–516

36 Mack, A. and Rock, I. (1998) *Inattentional Blindness*, MIT Press

37 Lamme, V.A. (2004) Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural Netw.* 17, 861–872

38 Schneider, W. and Shiffrin, R.M. (1977) Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev.* 84, 1–66

39 Shiffrin, R.M. and Schneider, W. (1977) Controlled and automatic human information processing: II. Perceptual learning, automaticity, attending and a general theory. *Psychol. Rev.* 84, 127–190

40 Shah, P. and Miyake, A. (1999) Models of working memory: An

introduction. In *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (Miyake, A. and Shah, P., eds), pp. 1–27, Cambridge University Press

41 Baars, B.J. and Franklin, S. (2003) How conscious experience and working memory interact. *Trends Cogn. Sci.* 7, 166–172

42 Baars, B.J. *et al*. (2003) Brain, conscious experience and the observing self. *Trends Neurosci.* 26, 671–675

43 Faraone, S.V. and Biederman, J. (1998) Neurobiology of attention-deficit hyperactivity disorder. *Biol. Psychiatry* 44, 951–958

44 Duffy, J.D. and Campbell, J.J. (2001) Regional prefrontal syndromes. In *The Frontal Lobes and Neuropsychiatric Illness* (Salloway, S. *et al*., eds), pp. 113–123, American Psychiatric Publishing

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

46 Pascual-Leone, A. and Walsh, V. (2001) Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292, 510–512

47 Edelman, G.M. and Tononi, G. (2000) *A Universe of Consciousness: How Matter Becomes Imagination*, Basic Books

48 Tononi, G. and Edelman, G.M. (2000) Schizophrenia and the mechanisms of conscious integration. *Brain Res. Brain Res. Rev.* 31, 391–400

49 Rumelhart, D.E. *et al*. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press

50 McClelland, J.L. and Rumelhart, D.E. (1981) An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychol. Rev.* 88, 375–407

51 McClelland, J.L. and Elman, J.L. (1986) The TRACE model of speech perception. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (McClelland, J.L. *et al*., eds), pp. 58–121, MIT Press

52 Mathis, W.D. and Mozer, M.C. (1996) Conscious and unconscious perception: A computational theory. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, pp. 324–328, Erlbaum

53 McGurk, H. and MacDonald, J. (1976) Hearing lips and seeing voices. *Nature* 264, 746–748

54 Perruchet, P. and Vinter, A. (2002) The self-organizing consciousness. *Behav. Brain Sci.* 25, 297–330

55 James, W. (1890) *The Principles of Psychology*, Henry Holt

56 Treisman, A. (1996) The binding problem. *Curr. Opin. Neurobiol.* 6, 171–178

57 Crick, F. and Koch, C. (1990) Towards a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275

58 Engel, A.K. and Singer, W. (2001) Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci.* 5, 16–25

59 Shadlen, M.N. and Movshon, J.A. (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* 24, 67–77, 111-125

60 Hummel, J.E. and Biederman, I. (1992) Dynamic binding in a neural network for shape recognition. *Psychol. Rev.* 99, 480–517

61 Hopfield, J.J. and Brody, C.D. (2001) What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proc. Natl. Acad. Sci. U. S. A.* 98, 1282–1287

62 Niebur, E. and Koch, C. (1994) A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *J. Comput. Neurosci.* 1, 141–158

63 Shastri, L. (1996) Temporal synchrony, dynamic bindings, and SHRUTI: a representational but non-classical model of reflexive reasoning. *Behav. Brain Sci.* 19, 331–337

64 von der Malsburg, C. (1981) *The Correlation Theory of Brain Function*, Internal Report (81-2), Max-Planck Institute for Biophysical Chemistry

65 von der Malsburg, C. and Schneider, W. (1986) A neural cocktail-party processor. *Biol. Cybern.* 54, 29–40

66 Fries, P. *et al*. (2002) Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. *J. Neurosci.* 22, 3739–3754

67 Lumer, E.D. *et al*. (1997) Neural dynamics in a model of the thalamocortical system. II. The role of neural synchrony tested through perturbations of spike timing. *Cereb. Cortex* 7, 228–236

68 Block, N. (1995) On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–287

69 Block, N. (2005) Two neural correlates of consciousness. *Trends Cogn. Sci.* 9, 46–52

70 Tong, F. (2003) Primary visual cortex and visual awareness. *Nat. Rev. Neurosci.* 4, 219–229

71 Clark, A. (1998) Magic words: How language augments human computation. In *Language and Thought: Interdisciplinary Themes* (Carruthers, P. and Boucher, J., eds), pp. 162–183, Cambridge University Press

72 Braver, T.S. and Cohen, J.D. (1999) Dopamine, cognitive control, and schizophrenia: the gating model. *Prog. Brain Res.* 121, 327–349

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

74 Lloyd, D.E. (2004) *Radiant Cool: A Novel Theory of Consciousness*, MIT Press

75 O'Reilly, R.C. and Munakata, Y. (2000) *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*, MIT Press