Modeling the Neural Basis of Cognitive Integration and Consciousness

Murray Shanahan and Dustin Connor

Department of Computing, Imperial College London,
180 Queen's Gate, London SW7 2AZ, United Kingdom
m.shanahan@imperial.c.uk

Abstract

This paper presents a number of models whose aim is to establish a computational basis for the hypothesis that conscious information processing in the brain is mediated by a mechanism of global broadcast. A possible role for this putative "global neuronal workspace" in achieving cognitive integration is mooted in the context of modular theories of mind, and an argument is advanced for its likely emergence within the sort of small-world brain network seemingly favoured by evolution. The paper concludes with some speculation on the relationship between life and consciousness as it could be.

Introduction

This article interweaves three strands of thinking in contemporary cognitive science. First, according to Baars (1988; 1997; 2002), the architecture of the mammalian brain comprises a number of parallel specialist processes (or modules) that compete and/or co-operate for access to a *global workspace*, in effect a mechanism for broadcasting information back to the whole cohort of specialists (Dehaene, *et al*., 2006; Shanahan, 2008a). The central claim of Baars's theory is that information processing which is local to the specialists is non-conscious and only broadcast information is consciously processed.

Second, advocates of modular theories of mind, despite the diversity of their views, are largely in agreement that some mechanism for transcending modular boundaries is a prerequisite for the highest levels of cognitive attainment (Fodor, 1983; 2002; Tooby & Cosmides, 1992; Mithen, 1996; Carruthers, 2002; 2006). This facilitates what Mithen (1996) calls *cognitive fluidity*, a capacity to integrate across distinct domains of expertise that promotes innovation and creativity (Wynn & Coolidge, 2004).

Third, it has been shown that cortical wiring in mammals exhibits the properties of a *small-world network* (Sporns & Zwi, 2004; Bassett & Bullmore, 2006). According to Striedter (2005), this is the consequence of evolutionary pressure to maintain communication between anatomically segregated regions in the face of an increasing neuron count, since this cannot go hand-in-hand with a proportional increase in connectivity

Drawing together these three themes, this article proposes that the long-range white matter connections that serve to keep down the average path length in large-scale cortical networks have been structured by evolution so as to develop into a global neuronal workspace (Dehaene & Naccache, 2001; Dehaene, *et al*., 2006; Shanahan, 2008a), which not only provides the integrative facility required to promote cognitive fluidity but is also a candidate for the neural substrate underlying consciousness (Shanahan & Baars, 2005). The argument draws on a variety of computer and robot models, and is capped with a short discussion of the relationship between consciousness as it is found in nature and consciousness as it could be.

Global workspace theory

Global workspace theory (Baars, 1988; 1997) is one of the most influential ideas in the burgeoning field of consciousness studies. Its basic tenets have been endorsed by respected philosophers (Dennett, 2001; Metzinger, 2003) and neuroscientists (Dehaene, *et al*. 1998; Dehaene & Naccache, 2001), and in cognitive psychology it has entered the undergraduate curriculum (Eysenck & Keane, 2005). Of course, the field is young and global workspace theory is open to future amendment or refutation. But it currently enjoys widespread support and a growing body of favourable evidence (Baars, 2002).

Central to the high-level, functional presentation of the theory is a computational architecture whose origins are in the blackboard systems of 1980s AI research. The architecture comprises a number of parallel, specialist processes and a *global workspace* (Fig. 1). The parallel specialists compete (and sometimes co-operate) to influence the global workspace, whose contents are broadcast back to the whole cohort of specialists, influencing them in turn. In operation, the architecture alternates between periods of competition and broadcast.

According to global workspace theory, the human brain instantiates the global workspace architecture, permitting a distinction to be drawn between conscious and non-conscious neural information processing. Information processing that takes place in the parallel specialists is non-conscious, while only information that is broadcast via the global workspace is consciously processed. Using the experimental paradigm of *contrastive analysis*, wherein closely matched conscious and non-conscious conditions are compared, this hypothesis can be tested empirically. Evidence to date using this method has been broadly supportive (Baars, 2002). Crucially, for

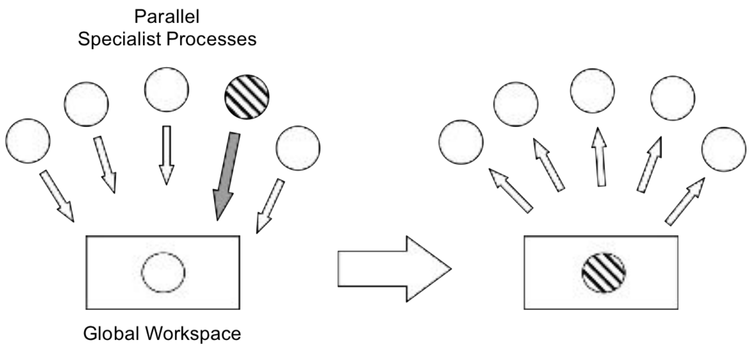

Fig. 1: The global workspace architecture. A set of parallel processes (shown as circles) compete for access to the global workspace (left). The winner (shown with hatched lines) influences the state of the global workspace, which is then broadcast back out to the whole cohort of processes (right). The resulting series of workspace states is the product of the repeated alternation between episodes of competition and broadcast.

contrastive analysis to be possible, both conscious and non-conscious processing must be capable of influencing behaviour. In the human case, introspective verbal report is typically taken as an index of conscious processing, while priming effects that occur in visual masking experiments are a good example of the influence of non-conscious processing (Breitmeyer & Öğmen, 2006).

Further support for global workspace theory can be garnered from its potential to bolster so-called modular theories of mind (Fodor, 1983; 2002; Tooby & Cosmides, 1992; Mithen, 1996; Carruthers, 2002; 2006). Modular theories of mind are challenged by the need for a mechanism that transcends modular boundaries, in order to implement what Fodor (2000) calls *informationally unencapsulated* cognitive processes, such as analogical reasoning, and to realise what Mithen (1996) calls *cognitive fluidity*. According to (Shanahan & Baars, 2005), the global workspace architecture incorporates just such a mechanism. Each parallel specialist process corresponds to a distinct module, and modular boundaries are transcended within the global workspace because the serial procession of states that unfolds there integrates the contributions of many of these parallel, specialists. Moreover, because the responsibility for determining the relevance of a potential contribution is not centralised but distributed among the specialists themselves, the resulting system is not vulnerable to the computational infeasibility arguments made by Fodor (1983; 2000).

One of the most pressing questions left open when global workspace theory is presented in functional terms is how the architecture maps onto the biological brain, and in particular what, in the brain, corresponds to the global workspace itself. A naive reading of the theory might attempt to associate the global workspace with a specific brain region, something reminiscent of the discredited notion of a Cartesian Theatre – "a place in the brain where it all comes together and consciousness happens" (Dennett, 1991). A more sophisticated understanding views the global workspace as an access-controlled, bandwidth-limited communications infrastructure that allows information to be distributed pan-cortically by means of global brain states. According to Dehaene and his colleagues, a *global neuronal workspace* of this sort is realised by the long-range cortico-cortical pathways of the cerebral white matter (Dehaene, *et al*., 1998; Dehaene & Naccache, 2001).

Modeling the global neuronal workspace

In order to realise a pan-cortical communications infrastructure and facilitate cognitive integration in accordance with the hypothesis of (Shanahan & Baars, 2005), the hypothesised global neuronal workspace should conform to the following four desiderata (Dehaene & Naccache, 2001; Shanahan, 2008a). 1) It should sustain reverberating patterns of activation over several tens of milliseconds. 2) It should disseminate (broadcast) patterns of activation throughout cortex, preserving the information inherent in their spatiotemporal structure. 3) It should be sensitive to new patterns of activation, and when overtaken by one only a trace should remain of any previous pattern. 4) Cortical populations should win the right to influence the pattern of activation in the workspace through competitive interaction.

One way to test the neurological plausibility of a global neuronal workspace conforming to these desiderata is to use the methods of computational neuroscience to build models of possible instantiations of the idea. In (Dehaene, *et al*., 2003) and (Dehaene & Changeux, 2005), a computer model is presented that simulates competitive access to a global neuronal workspace, emulating two well-known experimental phenomena, namely the attentional blink and inattentional blindness. But Dehaene's model does not simulate neuronal activity within the global workspace itself. In (Shanahan, 2008a), a complementary computer model is presented that simulates the (putative) global neuronal workspace itself in addition to a small number of cortical populations that compete to influence it (Fig. 2, left).

A schematic of the latter model is shown in Fig. 2 (right). Each box in the diagram represents a heterogenous population of over 1000 spiking neurons with conduction delays, implemented (in Matlab) using Izhikevich's (2003) equations. The global workspace comprises the five workspace nodes labeled W1 to W5, which serve to connect widely distributed regions of cortex. To keep the simulation manageable, only two such regions are included. Area W1 gives cortical

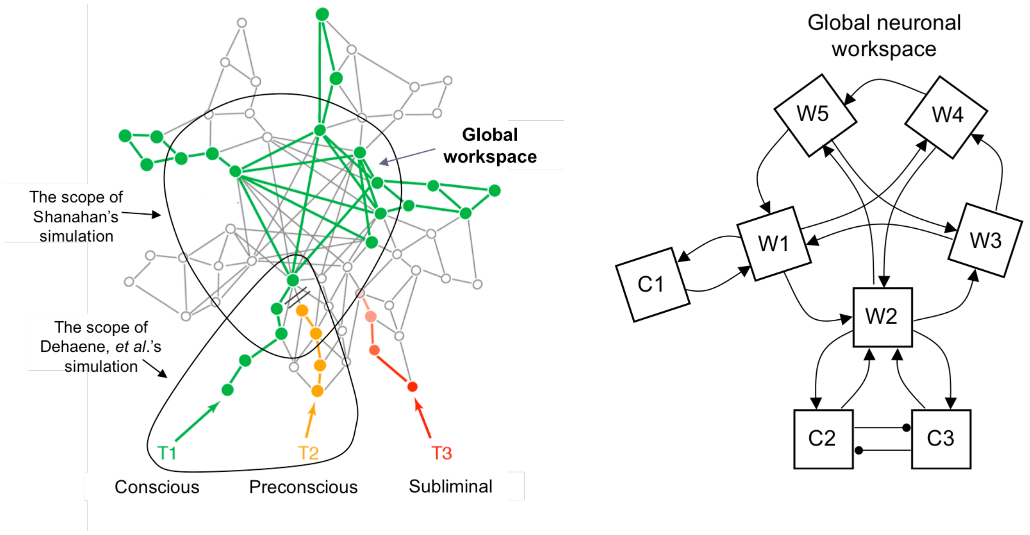
Global workspace
The scope of Shanahan's simulation
The scope of Dehaene, et al.'s simulation
T1
T2
T3
Conscious
Preconscious
Subliminal
Global neuronal workspace
W5
W4
W1
W3
C1
W2
C2
C3

Fig. 2: The global neuronal workspace (left) and its model (right). The brains of cognitively sophisticated animals can be thought of as instantiating the architecture of Fig. 1, with the long-range fibres of the cerebral white matter constituting a global neuronal workspace (left, adapted from Dehaene, *et al.* (2006)). The schematic on the right depicts the computer simulation described in (Shanahan, 2008a).

population C1 access to the workspace, while area W2 gives populations C2 and C3 access to the workspace. C2 and C3 are in a competitive relationship, mediated by local inhibitory connections as shown. All of the excitatory connections shown are focal and topographically organised, ensuring that the spatial structure of an activation pattern is preserved as it spreads out from a cortical population and into the workspace. The inhibitory connections between C2 and C3, on the other hand, are diffuse.

Not shown, but present in the model, are further diffuse inhibitory connections among the workspace nodes. Given the extensive recurrent connections between workspace nodes and the potential for feedback these provide, a suitable balance of excitation and inhibition is required to promote reverberation without preventing new paterns of activation from invading the workspace (Wang, 2001). Transitions from one workspace state to another are achieved thanks to the cortical populations C1 to C3. Each of these is trained, using a form of spike-timing dependent plasticity (STDP), to respond to the appearance of a certain pattern *Q* in the workspace by taking on an associated pattern *R*, which may then invade the workspace in turn. Suppose pattern *Q* is presently in the workspace. If C2 associates *Q* with *R* and C3 associates *Q* with *S* then a competition will ensue. If C2 wins the competition, the next workspace state will be *R*. This in turn may stimulate another cortical area to respond (C1 perhaps). Overall, the system alternates periods of broadcast with bursts of competition, and the workspace exhibits a procession of

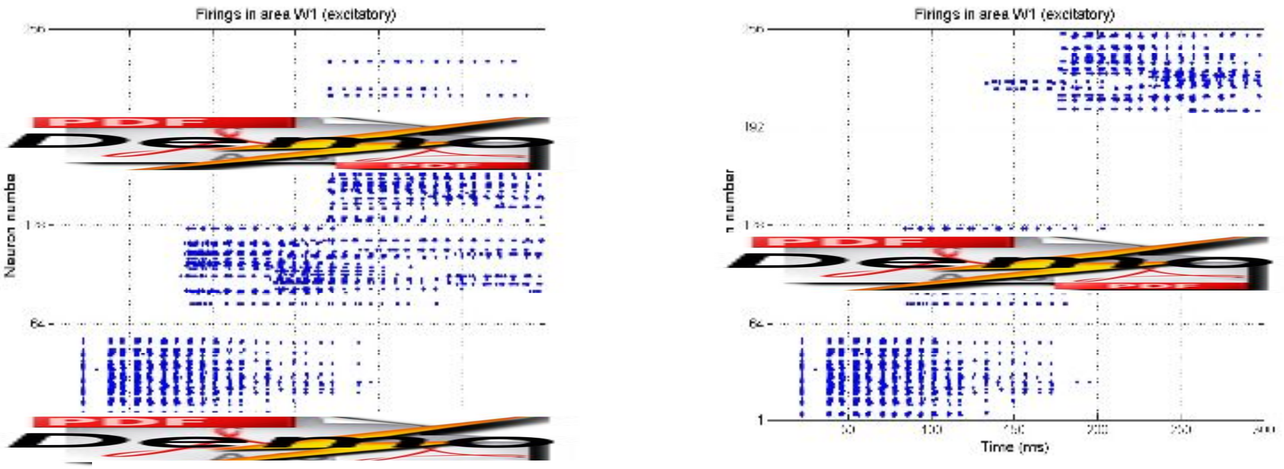

Fig. 3: Raster plots of neuron firings in two representative trials of the model of (Shanahan, 2008a). Both trials use the same network with identical synaptic weights. The difference is due to the competition between cortical popoulations C2 (influencing neurons 129–192) and C3 (influencing neurons 193–256), both of which respond equally strongly to activation in neurons 65 to 128, but with different associations. In the left-hand plot, C2 is the winner of the competition, shutting out its opponent by means of lateral inhibition, while in the right-hand plot the winner is C3.

broadcast states. Each of the components of the schematic in Fig. 2 (right) requires further internal structure to realise this behaviour. For full details the reader is referred to (Shanahan, 2008a) and (Shanahan, 2008b).

Fig. 3 shows raster plots of two representative trials of the simulation. For presentational purposes, the initial stimulus and the responses offered by C1 to C3 each activate a distinct set of contiguously numbered neurons. Firings in the excitatory neurons in workspace area W1 are shown. The other four workspace areas exhibit similar patterns, as we should expect if the workspace is operating effectively as a broadcast mechanism. In each trial, an initial stimulus is injected into the workspace at 20ms, which institutes a pattern of reverberating firing. C1 has an association with this particular pattern, and the pattern of firing it responds with begins to invade the workspace at around 80ms. This causes a surge of inhibition in the workspace thanks to which the original stimulus fades. By around 175ms almost no trace is left of it in either run.

At this point the two trials diverge. Areas C2 and C3 both have associations with the pattern of activation in the workspace, and a competition between them ensues. In the left-hand run C2 wins the competition, causing its response to take over the workspace, while in the right-hand run C3 is the victor. Note that in each case there is an outright winner, which prevents its rival from exercising any influence at all on the workspace. These trials were generated using the same network, with identical synaptic weights resulting from the same training run. The only source of difference between them is a small noise term added to the base current of each neuron. So taken together the two trials show that small differences at the level of individual neuron firings can result in qualitatively different sequences of workspace states at the macroscopic level (cf. Izhikevich & Edelman (2008)). A more complete description of the range of behaviours that can arise over multiple trials with differently trained networks can be found in Shanahan's papers (2008a; 2008b).

A workspace with stochastic wiring

The model of (Shanahan, 2008a) conforms to the desiderata set out earlier. But its neurological plausibility is compromised by the overly regular character of the workspace wiring. To address this shortcoming, work is ongoing to build and study global workspace models in which the long-range recurrent connections that promote reverberation and broadcast are established with a stochastic method that better reflects the statistical character of the evolutionary and developmental processes by which the brain's white matter pathways are formed.

In the new model, the 1280 excitatory workspace neurons, rather than being partitioned into five distinct sets as in the previous simulation, are arranged in a ring, which immediately induces a distance measure between any two neurons (Fig. 4, left). The workspace is then wired up by repeatedly forming circuits of connections. Each neuron in a circuit is selected randomly, subject to the constraint that no two neurons in the same circuit are allowed to be too close to each other, and the circuits are of variable length. Because each neuron is self-exciting via a circular route of connections, reverberating activation is promoted. But because recurrent connections cannot stimulate nearby neurons, the spatial organisation of a pattern of activation is preserved rather than smeared as it spreads throughout the workspace.

As with the previous model, excitatory influences in the workspace must be balanced with inhibitory connections, to ensure that reverberation is not so strong that it prevents new patterns from forming. In the stochastically-wired workspace, this is achieved using a second ring of 320 inhibitory neurons, concentric to the first. Each of these inhibitory neurons is excited locally, enabling it to detect patches of high firing, but has a widespread inhibitory effect on the workspace. The idea is to allow strong patterns of activation to damp rival workspace activity.

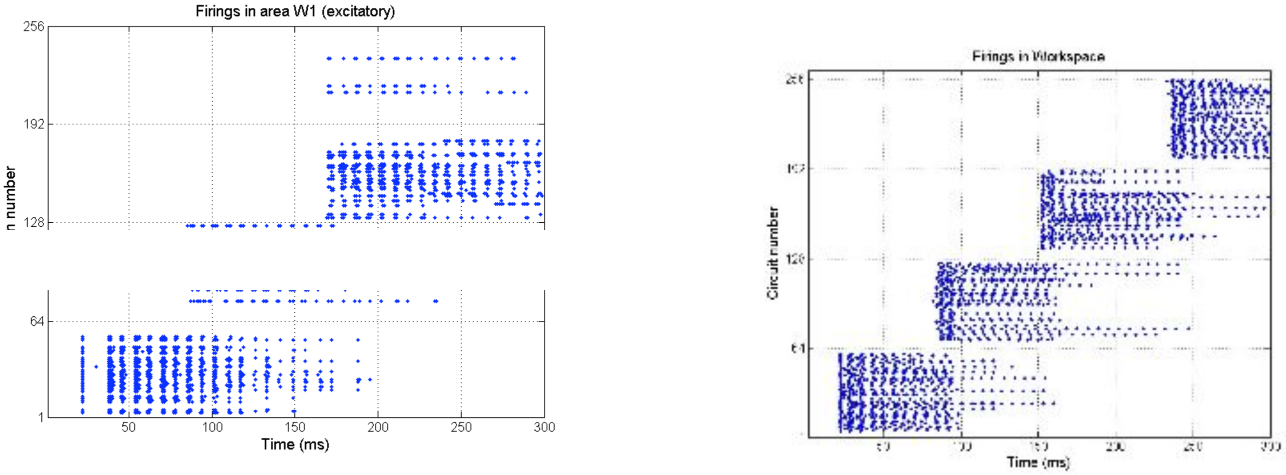

Fig. 4: A workspace with stochastic wiring. The workspace itself is a ring of neurons. Patterns of activation are broadcast (reverberate) around the ring via circuits of excitatory connections like the example shown on the left. Each circuit is wired up stochastically, but no two neurons in the same circuit are permitted to be close to eachother in the ring. Inhibitory neurons are locally excited but have diffuse influence (centre). The representative raster plot on the right shows that the workspace conforms to the desiderata.

Fig. 4 (right) shows a raster plot of a representative run in which a succession of four stimuli is delivered directly to the workspace. (The present model consists of the workspace only, and so far lacks the cortical populations of the previous model.) Each point in the plot represents that at least one neuron in the relevant circuit has fired. As the figure shows, the workspace maintains reverberation over several tens of milliseconds, and is susceptible to new patterns of activation which tend to push out their predecessors. In other words, the workspace conforms to three of the four desiderata proposed earlier, the fourth being inapplicable in the absence of cortical competition. Ongoing work aims systematically to map the range of model parameters and the space of possible network topologies that yield qualitatively equivalent behavioural characteristics.

Embodiment and cognitive architecture

The distinction between conscious and non-conscious processing that is the target of global workspace theory is only amenable to empirical investigation insofar as it impacts on outward behaviour. But the two computer models presented above are disembodied, closed systems. They must be embedded in a complete cognitive architecture if they are to stand as useful investigative tools. In (Shanahan, 2006), a cognitive architecture is presented that shows how a global workspace can be used in combination with an internally closed sensorimotor loop to realise a form of cognitively mediated action selection for a robot.

The central idea is that the internally closed sensorimotor loop permits the robot to rehearse trajectories through its sensorimotor space prior to enacting them (Hesslow, 2002). Rehearsed trajectories are evaluated, and the relative salience of the set of currently executable actions is modulated as a result – those initiating a trajectory whose outcome is associated with reward become more salient while those whose outcome is associated with aversion become less salient. Using a winner-takes-all strategy, the most promising action is selected and executed.

In the architecture of (Shanahan, 2006), the circuitry that makes up the inner sensorimotor loop takes in the global workspace itself (Fig. 5, left), and the series of rehearsed sensorimotor states unfolds within it. Hence these states are made available to the whole cohort of specialist networks that are attached to the workspace, enabling the trajectory of rehearsal to be determined by competition among those networks. Fig. 5 (right) presents the high-level schematic for a rationalised and extended version of the architecture. Internal sensorimotor activity, corresponding to that generated by the internally closed loop in Fig. 5 (left), results from mutual stimulation among motor and sensory areas, mediated by the global workspace. External sensory input causes activity in the sensory areas, which gives rise to activation in the workspace, from where it propagates to motor circuits. This stimulates a competition among motor areas to respond. During rehearsal, the resulting motor activity does not issue in overt behaviour, but instead gives rise to further, internally mediated stimulation of the sensory regions, completing the inner sensorimotor loop.

To date, the emphasis of our modeling work has been competition. Competitive access to the global workspace facilitates search through the sensorimotor space of a robot or animal because, as Fig. 3 shows, in cases where a sensorimotor state has multiple associations its successor in the workspace is non-deterministic. So revisiting a state can precipitate the rehearsal of an unexplored trajectory. But the hypothesis of the present paper is that the potential for co-operation among different networks might be equally

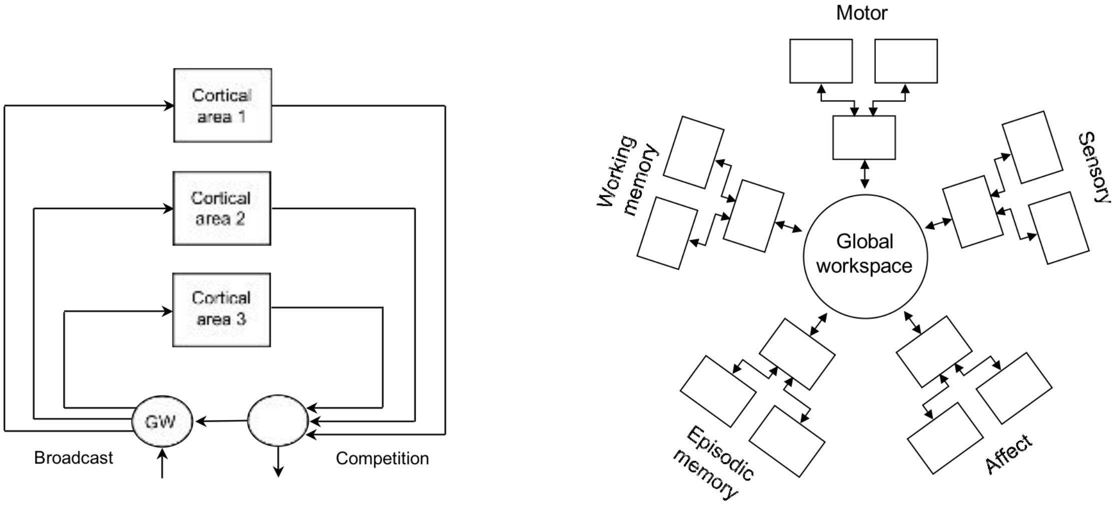

Fig. 5: Combining a global workspace with an inner sensorimotor loop (left), and a proposed rationalisation and extension of the architecture (right, cf. Fig. 1 of (Friston, 2003)). In both diagrams, information fans out from the global workspace into many distributed, parallel networks (broadcast) and funnelling back into it from those networks (competition). In the new architecture (right), five broad categories of functionally distinct networks are shown, each having a hierarchical structure. Co-operation, co-ordination, and competition among these networks is mediated by the global workspace, which best thought of as an access-controlled, bandwidth-limited communications infrastructure, rather than a functional component in its own right.

important. This is because co-operation may permit the adaptation and combination of elements from different parts of a learned repertoire of sensorimotor patterns, possibly enabling rehearsal even in a novel situation, such as that faced by an animal in the classic trap-tube test of causal understanding (Povinelli, 2000). To pass this test, an animal has to select the end from which to push a food item out of a transparent tube. The wrong choice results in the reward falling into a hole in its path which is visible to the animal. Some non-human animals, including chimpanzees and crows, are able to pass variants of this test, although there is no consensus among animal cognition researchers about how they do it (Seed, *et al*., 2006; Penn & Povinelli, 2007). A normal human adult, of course, is not unduly taxed this problem. Indeed, such capacity to innovate in the presence of novelty is often taken as a hallmark of human-level intelligence (Wynn & Coolidge, 2004).

According to the present hypothesis, the integrative facility of the global workspace supports the level of cognitive sophistication required to solve problems such as the trap-tube test. Networks encoding incompatible learned sensorimotor patterns are obliged to *compete* to influence the trajectory of rehearsal as it unfolds in the workspace. But where different networks encode compatible spatiotemporal patterns, they may be able to *co-operate*, allowing their respective influences to be blended together. Each specialist process may be thought of as encapsulating expertise in a particular micro-domain, such as dropping-things-in-holes or pushing-things-with-sticks. In effect, the global workspace promotes cognitive fluidity, permitting expertise in one micro-domain to be combined with expertise in another micro-domain (Shanahan & Baars, 2005). Our future work aims to explore this hypothesis with the aid of a large-scale spiking neuron implementation of the architecture of Fig. 5 (right), deployed to control a dextrous humanoid robot.

The emergence of a (small) world

Recent studies of neural connectivity lend further support to the hypothesis that cognitively proficient brains conform to the global workspace architecture. In particular, there is compelling evidence that human cortex constitutes a *small-world network* (Watts & Strogatz, 1998), which is a sparsely connected graph with a small mean path length and a large clustering coefficient. Consider a graph *G* comprising a set of nodes and edges. The *path length* between any pair of nodes in *G* is the number of edges in the shortest path between those nodes, and *G*'s *mean path length* is the path length averaged over every pair of nodes in *G*. The *clustering coefficient* of a node *P* in *G* is the fraction of the set of all possible edges between immediate neighbours of *P* that are actual edges, and the *clustering coefficient* of the whole graph *G* is the clustering coefficient averaged over the set of all nodes in *G*. Many naturally occurring networks have been shown to have small-world properties, but our concern is only with those that are found in the brain.

A typical small-world network comprises numerous densely interconnected local clusters that are connected to each other via a small number of so-called *hub nodes* but are otherwise isolated. If the hub nodes have many edges compared to the cluster nodes then such a network may also be *scale-free*, meaning that the probability of a random node having *k* edges conforms to a power law – it is proportional to $k^{-\lambda}$ for some λ. However, as we shall use the term, a node does not require a large number of edges to be designated a hub. A hub node may, for example, be the only node in cluster *A* that is connected to a node in cluster *B*, thus helping to confer the small-world property on the overall graph. (In graph-theoretic terms, such nodes have low degree but high betweenness centrality.)

Even without formal analysis, it is easy to see that the topology of the model in (Shanahan, 2008) (Fig. 2, right) leads to small-world connectivity. First, thanks to the connections to, from, and between workspace neurons (the hub nodes), the maximum shortest path length between any two cortical neurons is just 6, even with the addition of further cortical areas (2 hops to get to a workspace area, 2 more to traverse the workspace, plus 2 to get out of the workspace). Second, the dense connectivity within cortical areas entails a high clustering coefficient. Finally, although the network is not especially sparse with only C1 to C3 attached to the workspace, its sparseness increases rapidly with the addition of further cortical areas. Similar considerations apply with the stochastically wired workspace.

Using neuroanatomically established connectivity matrices, it has been shown that the cortices of cats and macaques enjoy small-world network properties (Hilgetag, *et al*., 2000; Sporns & Zwi, 2004). Moreover, several recent *in vivo* studies purport to establish similar results for human cortex. Using fMRI, Eguíluz, *et al*. (2005) revealed a network of functional brain connections that conform to the power law characteristic of a scale-free, small-world network. Also using fMRI, Achard, *et al*. (2006) confirmed this result and built a connectivity map of the cortical hub nodes underlying it. At the structural level, He, *et al*. (2007) supply a similar map by correlating measures of cortical thickness in different brain regions obtained by MRI.

The question of why evolution should favour neural networks with small-world properties naturally arises. A number of answers have been suggested (Bassett & Bullmore, 2006). Wiring cost is likely to be one major factor (Striedter, 2005; Wen & Chklovskii, 2006). If connectivity is maintained as brains increase in neuron count, then the quantity of wiring must increase too. Wiring is costly "due to metabolic energy required for maintenance and conduction, guidance mechanisms in development, conduction time delays and attenuation, and wiring volume" (Wen & Chklovskii, 2006, p.0617). But pressure to minimise wiring can lead to a network that is segregated into clusters (or modules). A small-world network compensates for this by allowing effective communication to be maintained between distant regions (Striedter, 2005). At the same time, pressure to minimise conduction delays may also lead to small-world properties, as well as the division of the brain into grey and white matter (Wen & Chklovskii, 2006).

In addition to their favourable wiring cost, small-world networks have been shown to possess information processing characteristics that make them especially well-suited to realising a global neuronal workspace. Specifically, Sporns, *et al* (2000) argue that small-world networks support high dynamical "complexity", according to a formal measure that assesses the co-existence in a network of functional specialisation and integration (Tononi, *et al*., 1998; Seth, *et al*., 2006). According to this measure, the complexity of a system X comprising n variables x_i is approximated by the function $C(X)$, given by

$$C(X) = H(X) - \sum H(x_i \mid X - x_i)$$

where $H(Y)$ is the entropy of a system Y and $H(y \mid Y)$ is the conditional entropy of y given Y. In essence, if a system has a low level of integration then values for $H(x_i \mid X)$ will be high, while if the system has a low degree of specialisation the value of $H(X)$ will be low. Using an evolutionary algorithm, Sporns, *et al*. searched a space of possible network topologies, selecting for networks with high $C(X)$. A typical network obtained after 2000 generations with this method had a mean path length comparable to that of an equivalent random graph, but a significantly higher clustering coefficient.

Intuitively, this result makes perfect sense. At a local level, the densely interconnected clusters of a small-world network are functionally segregated, while at a global level the connections between hub nodes ensure that the network's overall activity has widespread local influence. Moreover, it should be clear that a capacity to support high dynamical complexity in the sense quantified by $C(X)$ is a prerequisite for any neural network instantiation of the global workspace architecture, and that a network with small-world properties supplies the means to fulfil this prerequisite. The local specialists of the global workspace architecture can be realised by the highly interconnected, functionally segregated clusters of a small-world network, ensuring a high value for $H(X)$, while the global workspace itself is realisable by a web of hub-node-to-hub-node connections, promoting low values for $H(x_i \mid X)$.

Additional organisation over and above small-world topology is required for a network to conform to the desiderata set out earlier and realise the function of a global neuronal workspace. But only a relatively conservative set of modifications to the hub node connections of a sufficiently large small-world network may be needed for their integrative potential to be recruited to this role. Of course, once these modifications have been selected for, their cognitive advantages will ensure their perpetuation. But it is an intriguing thought that consciousness might initially have arisen only as a side-effect of the evolutionary pressure to keep wiring cost down, a constraint that applies across the phylogenetic scale from *C.elegans* upwards, but which ensures that the necessary infrastructure to support the distinction between conscious and non-conscious processing is already in place as neuron count goes up.

Consciousness as it could be

Artificial life, according to one of the field's founders, "can contribute to theoretical biology by locating *life-as-we-know-it* within the larger picture of *life-as-it-could-be*" (Langton, 1989, p.1). In a similar vein, the use of computer and robot models might aspire to contribute to cognitive science by situating consciousness as we know it within the larger picture of consciousness as it could be. No less interesting is the challenge of situating consciousness as it could be in relation to life as it could be. Indeed several authors argue for the deep continuity of life and mind: "life and mind share a set of basic organizational properties, and the organizational properties distinctive of mind are an enriched version of those fundamental to life" (Thompson, 2007, p. 128).

The argument for this position is roughly as follows. An organism perpetually constitutes its own identity through metabolic exchange of matter and energy with the environment so as to maintain the boundary between self and non-self. At the same time this "autopoietic" process brings forth a domain of concern, wherein features of the environment acquire significance according to their relevance to that organism's wellbeing and perpetuation. Moreover, an organism's need constantly to change in order simply to maintain its identity opens up what phenomenologists call a temporal and spatial "horizon" for that organism. For phenomenologists, such a "horizon of transcendence" is also a necessary feature of lived experience, motivating the conclusion that "certain existential structures of human life are an enriched version of those constitutive of all life" (Thompson, 2007, p.157).

Let's review the principles of organisation claimed in this paper to be fundamental to consciousness, and consider the extent to which they resonate with the thesis of deep continuity of life and mind. The global workspace architecture harnesses the power of massively parallel computation. The global workspace itself exhibits a *serial* procession of states, yet each state-to-state transition is the result of filtering and integrating the contributions of huge numbers of *parallel* computations. In essence, the architecture thereby distils *unity* out of *multiplicity*. This unity is achieved within the global workspace itself, which is both the source and sink of information in the fan and funnel model (Fig. 5, left). But it is also a locus of control, and the informatic singularity of the global workspace is inherently bound to the spatially localised body whose control is in question, the point of convergence of perception and action (Legrand, 2006). The *remit* of all the processes that are brought into unity by the global workspace is duly inherited from the body to which it is bound (Shanahan, 2005). Everything they do pertains to, or is *indexical* to, that body and its point of view.

In the natural world this remit in large part subserves metabolism, and is plausibly cast in terms of autopoiesis. But in the realm of the possible, of consciousness as it could be, metabolism is not a prerequisite for being a centre of concern, for possessing self-related purpose within a spatial and temporal "horizon of transcendence". In a properly embodied instantiation of the global workspace architecture, the identity of the conscious subject is underwritten by the common remit of a set of processes that pertain to the past, present, and future of the spatially localised body to which they are all indexically oriented (Fig. 5, right). In conclusion, however formidable the practical obstacles might be to creating a conscious artefact, the absence of metabolism presents no obvious theoretical obstacle. Perhaps the appeal of the deep continuity thesis is attenuated by this caveat.

References

Achard, S., Salvador, R., Whitcher, B., Suckling, J. & Bullmore, E. (2006). A Resilient, Low-Frequency, Small-World, Human Brain Functional Network with Highly-Connected Association Cortical Hubs. *Journal of Neuroscience* 26 (1), 63–72.

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

Baars, B.J. (1997). *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press.

Baars, B.J. (2002). The Conscious Access Hypothesis: Origins and Recent Evidence. *Trends in Cognitive Sciences* 6 (1), 47–52.

Bassett, D.S. & Bullmore, E. (2006). Small-World Brain Networks. *The Neuroscientist* 12 (6), 512–523.

Breitmeyer, B. & Öğmen, H. (2006). *Visual Masking: Time Slices Through Conscious and Unconscious Vision*. Oxford University Press.

Carruthers, P. (2002). The Cognitive Functions of Language. *Behavioral and Brain Sciences* 25 (6), 657–674.

Carruthers, P. (2006). *The Architecture of the Mind*. Oxford University Press.

Dehaene, S. & Naccache, L. (2001). Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework. *Cognition* 79, 1–37.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J. & Sergant, C. (2006). Conscious, Preconscious, and Subliminal Processing: A Testable Taxonomoy. *Trends in Cognitive Sciences* 10 (5), 204–211.

Dehaene, S., & Changeux, J.-P. (2005). Ongoing Spontaneous Activity Controls Access to Consciousness: a Neuronal Model for Inattentional Blindness. *PLoS Biology* 3 (5), e141.

Dehaene, S., Kerszberg, M. & Changeux, J.-P. (1998). A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks. *Proceedings of the National Academy of Science* 95, 14529–14534.

Dehaene, S., Sergent, C., & Changeux, J.-P. (2003). A Neuronal Network Model Linking Subjective Reports and Objective Physiological Data During Conscious Perception. *Proc. National Academy of Sciences* 100 (14), 8520–8525.

Dennett, D. (1991). *Consciousness Explained*. Penguin.

Dennett, D. (2001). Are We Explaining Consciousness Yet? *Cognition* 79, 221–237.

Eguíluz, V.M., Chialvo, D.R., Cecchi, G.A., Baliki, M. & Apkarian, A.V. (2005). Scale-Free Brain Functional Networks. *Physical Review Letters* 94, 018102.

Eysenck, M.W. & Keane, M.T. (2005). *Cognitive Psychology; A Student's Handbook*. Fifth edition. Psychology Press.

Fodor, J.A. (1983). *The Modularity of Mind*. MIT Press.

Fodor, J.A. (2000). *The Mind Doesn't Work That Way*. MIT Press.

Friston, K. (2003). Learning and Inference in the Brain. *Neural Networks* 16, 1325–1352.

He, Y., Chen, Z.J. & Evans, A.C. (2007). Small-World Anatomical Networks in the Human Brain Revealed by Cortical Thickness from MRI. *Cerebral Cortex* 17, 2407–2419.

Hesslow, G. (2002). Conscious Thought as Simulation of Behaviour and Perception. *Trends in Cognitive Sciences* 6 (6), 242–247.

Hilgetag, C.-C., Burns, G.A.P.C, O'Neill, M.A., Scannell, J.W. & Young, M.P. (2000). Anatomical Connectivity Defines the Organization of Clusters of Cortical Areas in the Macaque Monkey and the Cat. *Philosophical Transactions of the Royal Society B* 355, 91–110.

Izhikevich, E. M. (2003). Simple Model of Spiking Neurons. *IEEE Transactions on Neural Networks* 14, 1569–1572.

Izhikevich, E.M. & Edelman, G.M. (2008). Large-Scale Model of Mammalian Thalamocortical Systems. *Proc. National Academy of Sciences* 105 (9), 3593–3598.

Langton, C. (1989). Artificial Life. In C.Langton (Ed.), *Artificial Life*, Addison Wesley, pp. 1–47.

Legrand, D. (2006). The Bodily Self: The Sensori-motor Roots of Pre-reflexive Self-Consciousness. *Phenomenology and the Cognitive Sciences* 5, 89–118.

Metzinger, T. (2003). *Being No One: The Self-Model Theory of Subjectivity*. MIT Press.

Mithen, S. (1996). *The Prehistory of the Mind*. Thames & Hudson.

Penn, D.C. & Povinelli, D.J. (2007). Causal Cognition in Human and Nonhuman Animals: A Comparative, Critical Review. *Annual Review of Psychology* 58, 97–118.

Povinelli, D.J. (2000). *Folk Physics for Apes*. Oxford University Press.

Seed, A.M., Tebbich, S., Emery, N.J. & Clayton, N.S. (2006). Investigating Physical Cognition in Rooks, *Corvus frugilegus*. *Current Biology* 16, 697–701.

Seth, A.K., Izhikevich, E., Reeke, G.N. & Edelman, G.M. (2006). Theories and Measures of Consciousness: An Extended Framework. *Proc. National Academy of Science* 103 (28), 10799–10844.

Shanahan, M.P. (2005). Global Access, Embodiment, and the Conscious Subject. *Journal of Consciousness Studies* 12 (12), 46–66.

Shanahan, M. P. (2006). A Cognitive Architecture that Combines Internal Simulation with a Global Workspace. *Consciousness and Cognition* 15, 433–449.

Shanahan, M.P. (2008a). A Spiking Neuron Model of Cortical Broadcast and Competition. *Consciousness and Cognition* 17, 288–303.

Shanahan, M.P. (2008b). Supplementary Note on "A Spiking Neuron Model of Cortical Broadcast and Competition". *Consciousness and Cognition* 17, 304–306.

Shanahan, M.P. & Baars, B. (2005). Applying Global Workspace Theory to the Frame Problem. *Cognition* 98 (2), 157–176.

Sporns, O., Chialvo, D.R., Kaiser, M. & Hilgetag, C.-C. (2004). Organization, Development and Function of Complex Brain Networks. *Trends in Cognitive Sciences* 8 (9), 418–425.

Sporns, O., Tononi, G. & Edelman, G.M. (2000). Theoretical Neuroanatomy: Relating Anatomical and Functional Connectivity in Graphs and Cortical Connection Matrices. *Cerebral Cortex* 10, 127–141.

Sporns, O. & Zwi, J.D. (2004). The Small World of the Cerebral Cortex. *Neuroinformatics* 2 (2), 145–162.

Striedter, G.F. (2005). *Principles of Brain Evolution*. Sinaur Associates, Inc.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.

Tononi, G., Edelman, G.M. & Sporns, O. (1998). Complexity and Coherency: Integrating Information in the Brain. *Trends in Cognitive Sciences* 2 (12), 474–484.

Tooby, J., & Cosmides, L. (1992). The Psychological Foundations of Culture. In J. H. Barkow, L. Cosmides, & J.Tooby (Eds.), *The Adapted Mind*, Oxford University Press, pp. 19–136.

Watts, D.J. & Strogatz, S.H. (1998). Collective Dynamic of 'Small-world' Networks. *Nature* 393, 440–442.

Wang, X.-J. (2001). Synaptic Reverberation Underlying Mnemonic Persistent Activity. *Trends in Neuroscience* 24 (8), 455–463.

Wen, Q. & Chklovskii, D.B. (2006). Segregation of the Brain into Gray and White Matter: A Design Minimizing Conduction Delays. *PLoS Computational Biology* 1 (7): e78.

Wynn, T. & Coolidge, F.L. (2004). The Expert Neandertal Mind. *Journal of Human Evolution* 46, 467–487.