A Simulated Global Neuronal Workspace with Stochastic Wiring

Dustin Connor and Murray Shanahan

Department of Computing, Imperial College London, 180 Queen's Gate, London SW7 2AZ, England, UK
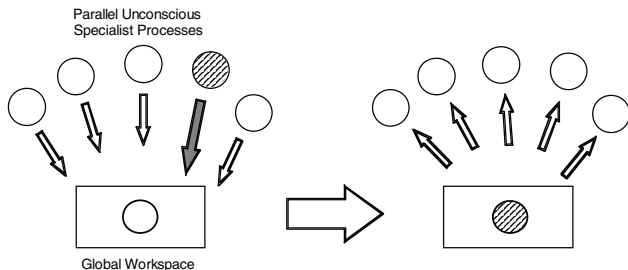dustin.connor02@imperial.ac.uk; mpsha@doc.ic.ac.uk

Abstract

It has been hypothesised that a global neuronal workspace underlies the brain's capacity to process information consciously. This paper describes a significant variation on a previously reported computer simulation of such a global neuronal workspace, in which competition and broadcast were realised through reverberating populations of spiking neurons. In the present model, unlike the previous one, the wiring of the workspace is not predefined but constructed in a stochastic fashion on the fly, which is far more biologically plausible. We demonstrate that the new model exhibits the same capacity for broadcast and competition as the previous one. We conclude with a short discussion of a brain-inspired cognitive architecture that incorporates a global workspace of the sort described.

1. Introduction

Global workspace theory is a widely favoured framework for the scientific study of consciousness. The theory is based on a parallel computational architecture that originated in the field of AI (Baars, 1988), and which, more recently, has found several applications in its original AI home (Franklin & Graesser, 1999; Shanahan, 2005). The architecture comprises a set of *parallel specialist processes* and a *global workspace* (Fig. 1). The parallel specialists compete (and sometimes co-operate) in a bid to win access to the global workspace. When one or more of them is successful in this bid, it gets to deposit a message in the global workspace, which is then broadcast back to the whole cohort of parallel specialists. So the overall pattern of i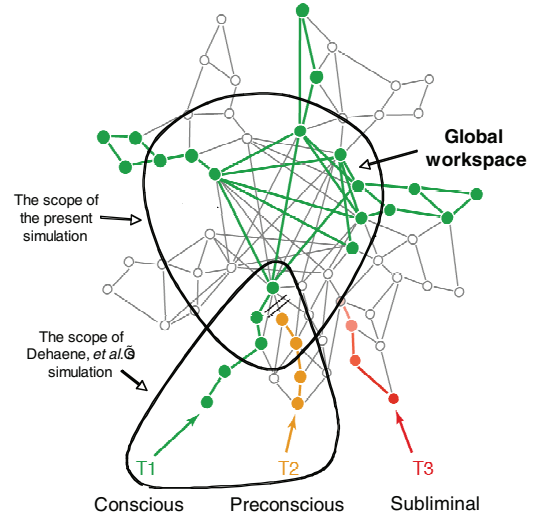nformation flow is one of alternation between *broadcast* and *competition*, resulting in a serial procession of workspace contents, the outcome of selecting among and integrating the results of many parallel computations.

From an empirical standpoint, global workspace theory hypothesises (1) that the mammalian brain instantiates such a global workspace architecture, and (2) that information processing confined to the parallel specialists is non-conscious while only information that passes through the workspace is consciously processed. The experimental paradigm of contrastive analysis (Baars, 1988) permits the empirical exploration of hypothesis (2) on the assumption that (1) is correct. But our present concern is with hypothesis (1). In particular, the question arises of how the brain might instantiate the global workspace architecture. What might be the neural substrate for the proposed global workspace?

One possibility is that the long-range corticocortical fibres of the cerebellar white matter constitute a *global neuronal*


Fig. 1: The Global workspace architecture. Information flow within the architecture is characterised by alternating periods of competition (left) and broadcast (right).

Fig. 2: The Global neuronal workspace and its simulation (adapted from Dehaene, *et al.*, 2006). The scope of Dehaene's model is competitive access to the workspace, and it does not include the workspace in full. The previous and current simulations incorporate a simplified form of competition, but also model the dynamics of broadcast.

workspace (Dehaene, *et al*., 1998; 2006; Dehaene & Naccache, 2001) – an infrastructure that allows for the dissemination (broadcast) of patterns of activation throughout the cortex, and influence over which is the outcome of cortical competition. This proposal has been explored by Dehaene and his colleagues in a number of computer simulations based on spiking neurons (Dehaene, *et al*., 2003; 2005). But their work has so far concentrated mainly on the mechanisms for access to the putative workspace and less on the dynamics and connectivity of the workspace itself. Accordingly, in a recent paper by one of the present authors, a complementary computer model of the workspace itself is presented (Shanahan, 2007) in which alternating periods of broadcast and competition are evident. (Fig. 2)

2. The Model

The idea for the current simulation is to use the existing working model (Shanahan, 2007, Fig. 3) as a springboard for further adjustments and ideas. In the previous model, there were cortical column structures attached to a central workspace 'hub'. This was split into five discrete areas, with various connections between them. The state of a workspace area at any one time represented a distinct pattern of activation, which was communicated in one-to-one connections amongst the workspace areas, and outwards to the cortical columns. These columns could then respond with a different pattern, which may or may not be integrated into the workspace and in turn broadcast around to the other columns.


Global neuronal workspace
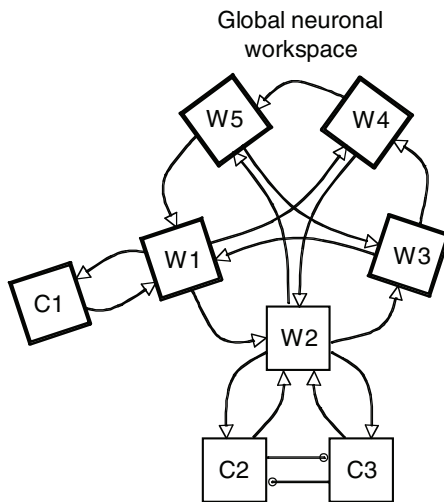
Fig. 3: Overall schematic of the previous model. The model comprises a number of interconnected workspace nodes (W1 to W5), plus three further cortical columns (C1 to C3).

This system displayed what have been theorised as the key features of a neuronal global workspace. There were alternating periods of competition between potential states for access to the workspace, and subsequent broadcast of the winner, followed by further competition. No processing took place in the workspace itself, just communication; however, a state which managed to get into the workspace propagated and was sustained over significant time periods. Although the workspace was sensitive to new states, it was important that the previous one did not continue to propagate after a new pattern started to exert an influence. To this end, a parallel inhibitory neuron system existed in the workspace. This would rarely degrade a single state, but would shut out an old state when a new one began to take over.

Considering these success, there are a few questions which can be raised about the structure of the workspace, bearing in mind our knowledge of real neuroanatomy. What does a single workspace area represent, and why would a neuronal workspace be split into discrete and identical areas at all? What rules govern the forming of connections between workspace neurons, and what features of the model's network are necessary for the desired behaviour to be observed? A potentially informative next step would be to generalise or at least change the network structure to some extent. This vein of investigation leads to the idea of spreading the five distinct areas into a single global workspace. Connections within this network would be more random and statistically governed, with some overarching rules to ensure some structure and points of comparison with the old model. This approach has the advantage of providing insight into the aforementioned questions while potentially providing more biologically plausible structures to study.

In this particular case, where we are making changes while trying to maintain the key features and behaviour, it makes sense to start by deciding what should stay constant between the two models. For easy comparison, it is useful to keep the same fundamental quantities: the same number of neurons in the workspace; similar ratio of inhibitory to excitatory neurons; same basic types of states propagating throughout; and the same model for the individual neurons themselves (Izhikevich, 2003). Workspace areas of the previous model represented different cortical localities where a state may need to be broadcast. Connections in the workspace were all *between* localities, not within them. This makes sense as it is often hypothesised that a neural correlate for the global workspace idea would probably be realised in long distance connections between various areas (Dehaene, 2006). To this end, the new model should consist solely of long distance connections within the workspace.
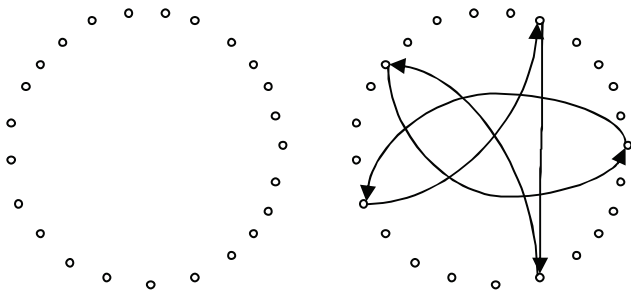
Fig. 4a (left): The neuron arrangement before being connected. **Fig. 4b (right)**: An example of a 'ring' of connections. Each neuron would be connected in a ring of this type.

Consider for a moment the meaning of 'long distance' in the new model. In the previous model, each area was distinctly separate from the others, and so connections between them were long distance. While the areas themselves were identical, they differed in that they were connected to different cortical columns. One can imagine that each of these columns represented functionally distinct areas in the brain. They have their own local connections to a subset of neurons which were in turn well connected to distant areas. A state which began in such a column but was strong enough to make its way to these long distance connections would then be broadcast to many other similarly connected neurons, and thus different columns. Without the distinction of different workspace areas, however, it is necessary to adopt a new measure of distance – with the aim of maintaining the same basic concept of functionally separate neurons communicating.

Now imagine an unconnected group of neurons – these are our workspace neurons. Instead of separating them into a few distinct groups, we can arrange them in a line, looping at the ends to make a circle (Fig. 4a). These neurons are not connected in a ring, simply arranged that way to emulate the spatial organisation of the cortical sheet (cf: Izhikevich, *et al*., 2004). If we assume that a cortical column could be attached to any given set of neighbouring neurons, then we can think of neighbouring neurons as functionally similar. The more neurons you move away from a certain place in the ring, the less functionality it will share with neurons at that location. A long distance connection can then be defined as any connection that is at least a certain number of neurons away in the ring. Following from the previous discussion, we have a rule for constructing the network: all connections are between neurons at least a certain distance away from each other in the ring. However, this is not yet sufficient to begin constructing the workspace.

In order to preserve the integrity of a state within the workspace, each of the areas in the previous model were identical, with one-to-one connections between them. One danger of statistically blurring the distinct workspace areas

into a single, large group of neurons is a blurring of the particular state as it propagates through the workspace. Some kind of structure is needed to guarantee reverberation, without the state itself changing or spreading across all possible patterns. The initial idea behind one-to-one connections between workspace areas is that different areas could mimic each other, providing the same information to multiple distinct places. That can be extended to the single neuron level in the model, where a particular neuron is mirrored in several places in the workspace. While this new model does not contain the distinct areas, it is still clearly made up of individual neurons in different functional regions. By having several neurons, which are distant from each other in the workspace and connected only amongst themselves, the integrity of the workspace state can be maintained while propagating it to different locations in the network.

These ideas provide the foundation for the rules of constructing the workspace network. It should be as randomly governed as possible so that there are not any distinct 'sections' in the workspace. All connections will be between neurons of a predefined distance away, and these connections exist between groups of matched neurons. There are no connections between neurons that are not of the same group. The simplest realisation of these rules involves making one-way 'rings' of connections for each neuron, with no overlapping connections between the rings (Fig. 4b). Note that this is not the only way of reconciling such rules, and a great deal of further work can be done simply in exploring various different network models.

The details provided give the desired network. However a key feature is missing in order to replicate the behaviour we expect of a global workspace: inhibition. A global workspace should suppress previous states in favour of new input. The previous model had general inhibition from each workspace area to some of the others. This inhibition was modulated such that there was only moderate decay for activation characteristic of a single state, but activation of a second state gave rise to widespread inhibition. This ensured that only a recent, strongly driven state would last, while a previous remnant would disappear. Reasonably extending this to the new model, an inhibitory ring of roughly one fifth the size of the excitatory ring can be linked such that each inhibitory neuron takes input from a wide collection of nearby excitatory neurons, and then inhibits a random sample of the workspace (Fig. 5).

Unlike the previous model, the focus here is only on the propagation, reverberation and competition of states within the workspace itself. As such, no cortical columns are modelled explicitly, though they could hypothetically be connected to any given group of neighbouring neurons in the newly modelled workspace.
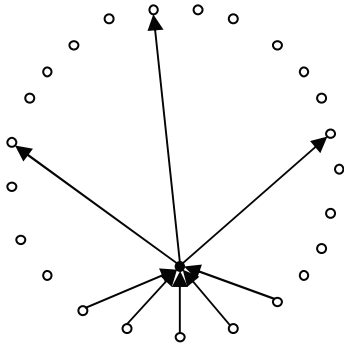
Fig. 5: The inputs and outputs of a typical inhibitory neuron. Inputs are locally organised, while outputs are randomly assigned.

2.1 Parameters

The new model consists of roughly 1280 excitatory neurons and 256 inhibitory neurons connected as explained in the previous section. Variation in the number of neurons arises when, in the construction of the network, a few neurons end up with no possible connections to other neurons within the outlined rules. These neurons are subsequently removed from the model to simplify the process. In an average test run of the model, only 2 or 3 such 'orphans' are found.

The distance parameter used to define the minimum length of a connection was 192 neurons. This was chosen to give ring sizes of 3-6 neurons, roughly equivalent to the five identical areas in the previous model. Consequently, though each setup can differ, on average there are 256 of these rings, with very little variance (±2 rings). It should be noted that in order to maintain an overall 'small-world' network, as described by research on the subject (Sporns 2004), the distance parameter would have to scale up with the size of the model. Issues of scaling any such computer model up are clearly relevant as neuron population numbers are far fewer than systems found in the actual human brain. The hope, of course, is that some scale-independent insights can come of models with a sufficiently large (though clearly not large enough to directly compare, due to computing limits) sample size.

2.2 Neuron Model

Individual neurons were simulated using Izhikevich's "simple model" of spiking behaviour (Izhikevich, 2003; 2007). This model is able to generate a large range of empirically accurate spiking behaviours, like the Hodgkin-Huxley equations, while being much easier to compute with. It is thus well suited to a large-scale, biologically plausible simulation. Moreover, the behaviour of the model is governed by four parameters (a, b, c, and d in Eqns. (1) – (3) below), which can be varied to emulate the signalling

properties of a wide variety of known neuron types. The model is defined by the following three equations:

$$\dot{v} = 0.04v^2 + 5v + 140 - u + I \qquad (1)$$

$$\dot{u} = a(bv - u) \qquad (2)$$

$$\text{if } v \geq 30 \text{then} \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \qquad (3)$$

where v is the neuron's membrane potential, I is its input current, and u is a variable that regulates the recovery time of the neuron after spiking. Eqn. (3) describes the way the neuron is reset after spiking, which is assumed to occur when its membrane potential reaches 30mV.

The values of the four parameters a, b, c, and d were lifted from (Izhikevich, 2003). For excitatory neurons these were $a = 0.02$, $b = 0.2$, $c = -65+16r^2$, and $d = 8-6r^2$, where r is a uniformly distributed random variable in the interval [0,1]. For inhibitory neurons, the values used were $a = 0.02+0.08r$, $b = 0.25-0.05r$, $c = -65$, and $d = 2$, with r as above. The random variable r introduces a degree of variation into each population. For example, in the excitatory case, if $r = 0$ we get the regular spiking behaviour shown on the left of Fig. 6, while if $r = 1$ we get the chattering behaviour shown on the right of the figure.

Consider a time t and a neuron i, and let F be the set of all neurons j that fired at time t–d where d is the conductance delay from neuron j to i. Then the input current I for neuron i at time t is given by:

$$I(t) = I_b + \sum_{j \in \Phi} S_{i,j} F \qquad (4)$$

where I_b is the base current, $S_{i,j}$ is the synaptic weight of the connection from neuron j to i, and F is a scaling factor whose value depends on the type of population to which i and j respectively belong (eg: workspace area, lateral inhibitory pool, etc). The scaling factors of the model are 60, 3.0, and 33.5 for workspace to workspace, workspace
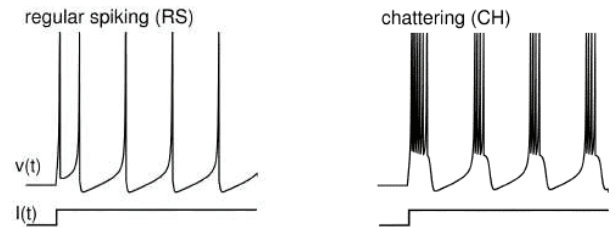
Fig. 6: Varieties of excitatory neurons using Izhikevich's simple model (from Izhikevich, 2003).

to inhibitory, and inhibitory to workspace connections respectively. Likewise, the conductance delays amongst workspace neurons and from inhibitory to workspace neurons are between 5 and 6ms. Delays from workspace to inhibitory neurons are 2ms. Scaling factors for pathways within the workspace are significantly higher than those to and from inhibitory neurons to compensate for the correspondingly smaller number of connections per neuron for those pathways.
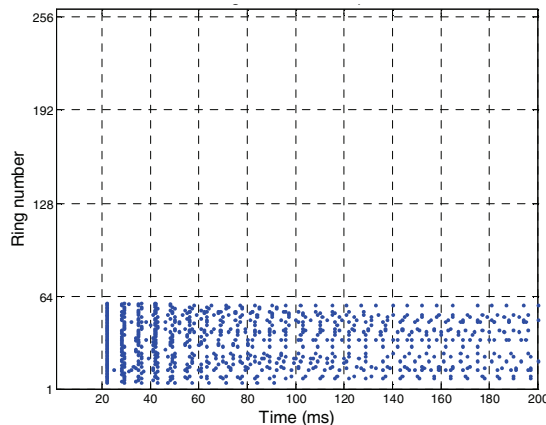
3. Experimental Results

Each experiment with the model consisted of applying stimuli to predetermined areas of the workspace at different times. These stimuli are simulated by a single pulse of 25mA to a particular set of rings in the given area, where different sets of rings correspond to different stimuli. In order to investigate the behaviour of this potential global workspace, the timing and origin in the network of the stimuli were varied across different trials.

Reverberation. A trial with a single stimulus shows considerable reverberation of a state. Although this state is not fully maintained indefinitely, it remains largely intact over hundreds of milliseconds (Fig. 7a). In addition, the workspace is equally sensitive to such a stimulus originating in any point of the network.

Competition. Adding a second stimulus at a later time shows that the system is sensitive to new input. The new state largely dominates the previous one, though there are some remnants. This can be repeated with subsequent different states, and old states return if re-stimulated (Fig. 7b).

The levels of both reverberation and competition depend closely on the parameters set for the strength of connections between neurons. As one might expect, stronger excitatory connections allow for more reverberation, but limit the ability of a new state to shut down an old one. Likewise, greater influence from the inhibitory population quickly shuts down old states in competitive circumstances, but also prevents any given state from lasting long.

Masking. One unpredicted feature of the new model becomes apparent when varying the timing of stimulus input. The workspace exhibits short windows of time when it is not sensitive to new stimulus input. These correspond to moments of high inhibition due to a recent stimulus's arrival. As a result, the model can seem temporarily 'blind', and miss a stimulus entirely (indistinguishable from Fig 7a). This phenomenon could be related to 'masking' effects noted in studies of human perception (Breitmeyer & Öğmen, 2006); however, much further investigation is necessary before such a claim is verified.

4. Discussion

This stochastically constructed network model displays many of the theorised behaviours necessary for a neuronal global workspace. It opens the door for further alterations to the structure of workspace networks, and brings to light some key features deserving of examination.

While studying this model it becomes clear that the timing of signals, dependent on both the conduction delays and the fundamental structure of the network, is important. Waves of inhibition, responsible for the takeover of new states as well as the gradual decay of old ones, depend on the synchronicity of local excitatory signals. Unpredicted events, such as 'masking', point to the need for a close
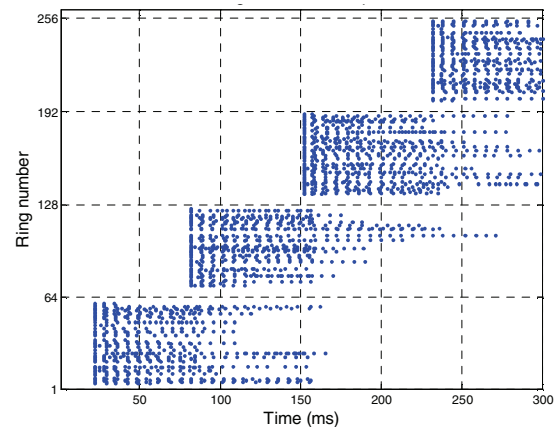
Fig. 7a (left): A typical trial with one stimulus at 20ms. Even with degradation, about half of the relevant signals for the state persist after 200ms. Notably, identical activation is seen when a second stimulus is presented in one of a couple 'masked' windows of time, such as 10ms after the first stimulus. **Fig. 7b (right)**: As different stimuli enter the workspace, previous states are shut out while new states begin to reverberate. This particular pattern of stimuli is comparable to a trial of the previous model.

analysis of the network's construction in order to decide whether these are simply artefacts of incorrect choices, or real effects to be considered.

Later iterations of this model could benefit from network constructions that are less rigid. The current 'ring' model allows information to travel only one-way and on one path in any given ring. This produces characteristic cycle times which will likely disappear in a slightly more integrated network.

In addition, it would be interesting to see how such a model works with cortical columns as part of a larger neuronal structure, rather than in a vacuum. The stochastic and single-path nature of each ring causes the information flow on a local level to be disorganised. As such, cortical columns may have to be temporally insensitive, so as to catch a more general picture of the state of the workspace rather than a snapshot. Alternatively, and perhaps more likely, the network would need more organisation than is implied by this model.

One reason why global workspace theory is of especial interest to AI researchers is that it posits an essential link between consciousness and cognition (Baars, 1988; Shanahan & Baars, 2005). Human cognitive prowess is surely possible thanks to the brain's ability to harness the massive parallelism inherent in its 100 billion neurons, and the global workspace architecture, with its ability to distil a serial stream of states that integrates the contributions of many parallel processes, suggests how this may be achieved. For example, the global workspace architecture is one possible way to tackle the problem, highlighted by Fodor, of supplying a computational account of the "informationally unencapsulated" cognitive processes, such as reasoning by analogy, or indeed of quickly assembling an effective response to a novel situation (Shanahan & Baars, 2005).

In (Shanahan, 2005), a brain-inspired cognitive architecture is described which has enabled some preliminary exploration of the engineering potential of a global workspace from a robotics perspective. The full architecture in that paper incorporates both a global workspace and an internally closed sensorimotor loop, which enables a robot to carry out an off-line simulation of the effects of its actions prior to actually carrying them out (Cotterill, 1998; Hesslow, 2002). The results of this process then inform the robot's choice of action. However, the global workspace used in the reported implementation of the architecture is built from simple neurons with no temporal dynamics. An interesting line of future work would be to re-implement this brain-inspired cognitive architecture using the global neuronal workspace described here, and to attempt to use it to control a dynamically interesting robot, such as that described in (Holland, 2006).

References

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

Breitmeyer, B. & Öğmen, H. (2006) *Visual Masking.* Oxford University Press

Cotterill, R. (1998). *Enchanted Looms: Conscious Networks in Brains and Computers.* Cambridge University Press.

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

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

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

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

Franklin, S. & Graesser, A. (1999). A Software Agent Model of Consciousness. *Consciousness and Cognition 8*, 285–301.

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

Holland, O. & Knight, R. (2006). The Anthropomimetic Principle. Proc. AISB Symposium on Biologically Inspired Robots, Bristol, Vol. II pp. 115–122

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

Izhikevich, E.M., Gally, J.A. & Edelman, G.M. (2004). Spike-Timing Dynamics of Neuronal Groups. *Cerebral Cortex 14*, 933–944.

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

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

Shanahan, M. A spiking neuron model of cortical broadcast and competition, *Consciousness and Cognition* (2007), doi:10.1016/j.concog.2006.12.005

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