To appear in *Consciousness and Cognition*

A Spiking Neuron Model of Cortical Broadcast and Competition

Murray Shanahan

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

December 2006

Abstract

This paper presents a computer model of cortical broadcast and competition based on spiking neurons and inspired by the hypothesis of a global neuronal workspace underlying conscious information processing in the human brain. In the model, the hypothesised workspace is realised by a collection of recurrently inter-connected regions capable of sustaining and disseminating a reverberating spatial pattern of activation. At the same time, the workspace remains susceptible to new patterns arriving from outlying cortical populations. Competition among these cortical populations for influence on the workspace is effected by a combination of mutual inhibition and top-down amplification.

Keywords: Consciousness, global workspace theory, global neuronal workspace, cortical competition, reverberating networks

1 Introduction

Global workspace theory has been highly influential among both philosophers and scientists interested in understanding consciousness (Baars, 1988; 1997; 2002). But the theory is commonly expressed in somewhat abstract terms, and it remains an open question how the architecture underlying the theory might be mapped onto the brain. According to one proposal, long-range cortico-cortical pathways realise a "global neuronal workspace" which enables a set of spatially distributed neural circuits to enter into a coherent, self-sustaining state during conscious episodes (Dehaene & Naccache, 2001).

One way to render such a hypothesis more concrete is to build and evaluate biologically realistic computer models of the neural circuitry that might realise the mechanisms proposed. Accordingly, models of various aspects of the hypothesised global neuronal workspace have been constructed by Dehaene and his colleagues (Dehaene, *et al.*, 1998; Dehaene, *et al.*, 2003; Dehaene & Changeaux, 2005). Continuing in this vein, the present paper describes a computer simulation of the hypothesised global neuronal workspace that incorporates mechanisms for both competitive access and broadcast, and in which a succession of distinct workspace states is exhibited.

In what follows, it will be assumed that cortical columns (or "modules") are a basic unit of neural processing (Mountcastle, 1997). According to known neuroanatomy, a portion of the neurons that comprise any given cortical column will connect it to distant cortical sites via the cerebellar white matter. These connections are likely to include direct cortico-cortical projections through bundles of association fibres, such as the arcuate fasciculus and the occipito-frontal fasciculi (Wakana, *et al.*, 2004), as well as indirect cortico-thalamo-cortical pathways mediated by what Sherman & Guillery (2002) call higher-order thalamic relays. The model presented here rests on the hypothesis that within certain cortical columns, called *workspace nodes*, a subset of such neurons exists that facilitates the flow of information to and from a global neuronal workspace, while the workspace itself is nothing more than the total set of such nodes plus the long-range pathways interconnecting them.

There is good evidence that cortical wiring, with its dense local connections and sparser long-range projections, enjoys the properties of a "small world" network (Sporns & Zwi, 2004). In theory, such an arrangement permits any given

cortical column to exert an influence over any other given column via a shortest path comprising only a handful of intermediate connections. According to the present proposal, workspace nodes can be thought of as so-called "hub nodes" in a large-scale, small-world cortical network, since their role is to link numerous local clusters of neurons to distant neural clusters via long-range connections into other hub nodes.

The present model comprises five workspace nodes and three further cortical columns including several populations of inhibitory and excitatory neurons. Reverberating patterns of activation are maintained in the workspace over several tens of milliseconds thanks to a balance of recurrent excitatory and inhibitory pathways between workspace nodes (Amit & Brunel, 1997; Wang, 2001). Competition for access to the workspace is governed by a combination of mutual lateral inhibition and top-down amplification, in a simplified version of the circuit used in (Dehaene, *et al*., 2003) and (Dehaene & Changeaux, 2005). Individual neurons are simulated using Izhikevich's "simple model" of a spiking neuron, which facilitates the efficient simulation of heterogeneous neural populations with biologically realistic behaviours (Izhikevich, 2003; 2007).

The paper is organised as follows. The next section supplies a short overview of global workspace theory, in which a number of guiding principles for the operation of the hypothesised global neuronal workspace are set out. The computer simulation is then presented, in terms of both its high-level architecture and the low-level neuron model deployed. The results of experiments with the simulation are then reported, with the behaviour of a single trial described in detail, and the outcome of a series of 36 trials summarised.

2 A Short Overview of Global Workspace Theory

Global workspace theory posits an empirical distinction between conscious and non-conscious neural information processing based on the hypothesis that the brain instantiates the architectural blueprint sketched in Fig. 1 (Baars, 1988; 1997). The architecture comprises a set of parallel specialist processes which compete for access to a global workspace. The process (or coalition of processes) that wins access gets to deposit a message in the global workspace, causing the message to be broadcast back to the entire cohort of parallel specialists. As Fig. 1 shows, the global workspace exhibits a serial procession of states. Yet the transition from one state to the next is the result of selecting from and combining

many parallel computations. As such, it has the potential to marshal the brain's massively parallel resources and orchestrate a unified, coherent response to the ongoing situation for an organism (Shanahan & Baars, 2005).

In the context of the global workspace architecture it is possible to posit the following distinction. Information processing carried out locally by the parallel specialists is non-conscious, and only information that is broadcast to the entire cohort is consciously processed. The validity of this distinction can be empirically tested using the experimental paradigm of contrastive analysis, wherein closely matched conscious and non-conscious conditions are compared (Baars, 1988), something made possible thanks to phenomena such as visual masking (Dehaene, *et al.*, 2001). Recent evidence acquired in this way is largely favourable to the global workspace idea, and the broad terms of the theory have attracted widespread approval (Baars, 2002).

However, our current level of understanding leaves many theoretical and empirical questions open. Not least among these is the question of exactly how the brain might instantiate the global workspace architecture. On a naïve reading, the above characterisation of the global workspace suggests a functionally and anatomically distinct entity, something akin to the Cartesian theatre discredited by Dennett – a "place in the brain where everything comes together and consciousness happens" (Dennett, 1991). But the most plausible way to map the architecture-level description onto actual brain mechanisms is to consider the workspace as a brain-scale "communications infrastructure" realised through a network of interconnected nodes distributed throughout the central nervous system. Thanks to this communications infrastructure, the activity in a single,

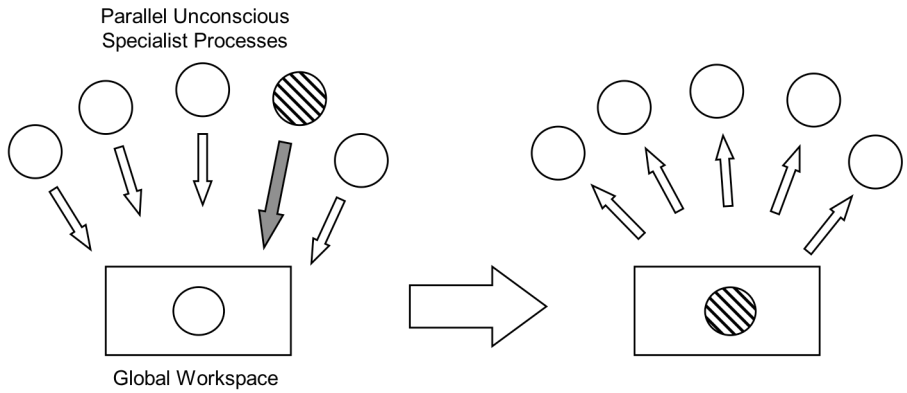

Fig. 1: The Global workspace architecture. Information flow within the architecture is characterised by alternating periods of competition (left) and broadcast (right).

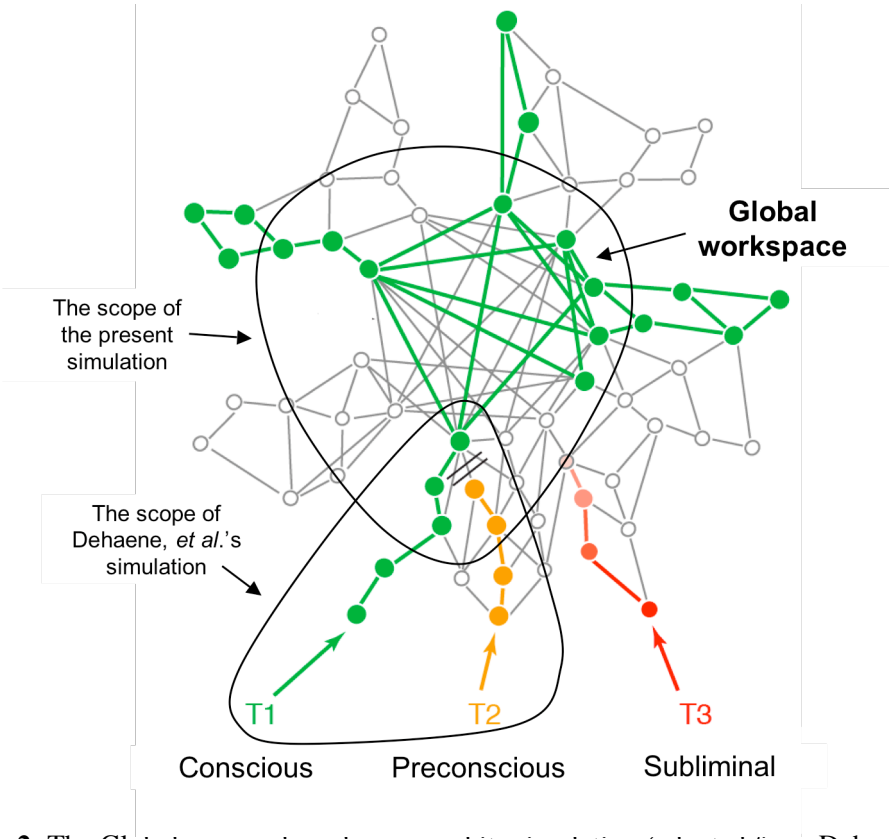

Fig. 2: The Global neuronal workspace and its simulation (adapted from Dehaene, *et al.*, 2006). The scope of Dehaene's model is competitive access to the workspace, and it does not include the workspace in full. The present simulation incorporates a simplified form of competition, but it also models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

localised cortical population can exercise a widespread, systematic influence on the activity of multiple distant cortical regions.

The global neuronal workspace proposed by Dehaene and Naccache (2001) conforms to this prescription (Fig. 2). They identify five classes of neural circuit that should, according to theoretical considerations, participate in a conscious state, namely high-level perceptual and motor systems, evaluative and attentional mechanisms, and long-term memory. They hypothesise the existence of "workspace neurons" linking these circuits together via long-range cortico-cortical fibres, and point to evidence from monkey studies of suitable pathways interconnecting many of the brain regions most closely associated with these functions, including the dorsolateral prefrontal, premotor, and anterior cingulate cortices, as well as various sub-cortical structures.

The hypothesis that this is a plausible anatomical substrate for the global workspace architecture gains support if a biologically realistic computer

simulation can be built that exhibits the sort of information flow the architecture requires. Two such simulations have been built by Dehaene and his colleagues. In the simulation of (Dehaene, *et al*., 1998), the workspace is modelled as a pool of neurons whose state is influenced by several outlying cortical processes, and which in turn has either an inhibitory or excitatory effect on those processes, realising a form of cortical selection. The focus of the more recent work reported in (Dehaene, *et al*., 2003) and (Dehaene, *et al*., 2005) is a computer model of competitive access to the global workspace. In this model, the flow of information is mostly into the workspace, and the only influence of the workspace areas on outlying cortical populations is top-down amplification. The present paper can be thought of as complementing the work of Dehaene, *et al*. Its contribution is to supply an explicit model of the mechanism and dynamics of broadcast, something which is not present in their simulation, and to demonstrate a two-way flow of spatial patterns both into and out of the modelled workspace. The relationship between Dehaene, *et al*.'s more recent simulation and the one reported here is illustrated by the two bubbles in Fig. 2.

To fulfil the cognitive function accorded to it by the high-level theory, the following four principles governing the operation of the hypothesised global neuronal workspace are proposed (Dehaene & Naccache, 2001).

- The workspace *sustains* patterns of activation over several tens of milliseconds.
- The workspace *disseminates* (broadcasts) patterns of activation throughout cortex, preserving the information inherent in their spatiotemporal structure.
- The workspace is *sensitive* to new patterns of activation, and when it is overtaken by them only a trace remains of any previous pattern.
- Cortical populations win the right to influence the pattern of activation in the workspace through *competitive* interaction.

The challenge now is to devise a detailed model of the global neuronal workspace which is compatible with the current state of neuroscientific knowledge, and to demonstrate that a computer simulation of the model can be built whose behaviour conforms to these four principles.

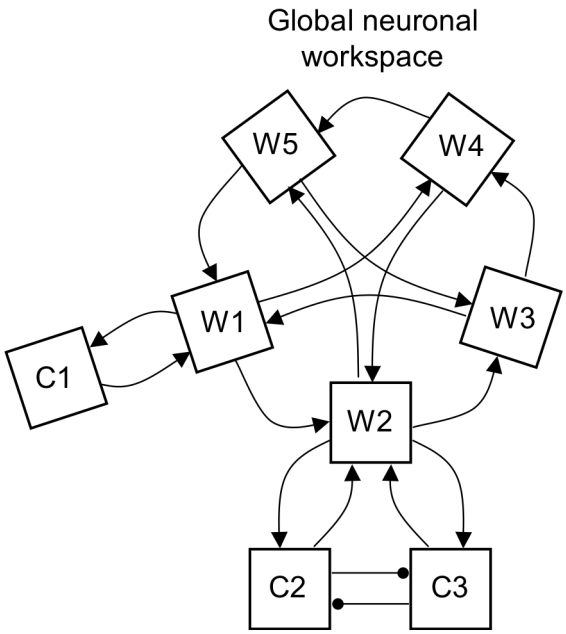

Fig. 3: Overall schematic of the model. The model comprises a number of interconnected workspace nodes (W1 to W5), plus three further cortical columns (C1 to C3). Columns C2 and C3 are near each other, and are mutually inhibiting.

3 The Computer Model

An overall schematic for the model is given in Fig. 3. The global workspace itself consists of five nodes (W1 to W5), each of which comprises a population of 256 excitatory and 64 inhibitory neurons. The workspace nodes are interconnected in such a way that activity in one node quickly spreads into the others, effecting a form of broadcast. The recurrent interconnections among the workspace areas promote reverberation, which has been used successfully to model various aspects of working memory (Compte, *et al.*, 2000; Deco & Rolls, 2003; Constantinides & Wang, 2004). The model also includes three further cortical columns (C1 to C3) capable of influencing the pattern of activation in the workspace, while the state of the workspace in turn influences the patterns of activation in those columns. Two of the columns (C2 and C3) compete for access to the same workspace node (W2).

A more complete computer model of the global neuronal workspace would consist of hundreds of workspace nodes and many hundreds of cortical columns. But to run such a model is at present computationally infeasible unless the columns and nodes are themselves idealised as a trivially small number of neurons. In the present simulation, each cortical column is modelled as a map of

approximately one thousand neurons, so that it can exhibit a spatiotemporally organised pattern of activation, while the number of workspace nodes and other columns is kept manageably small. With a realistically large number of workspace nodes, connections would be established on a statistical basis. To approximate this in the present model, with just five workspace nodes, each node is connected to two of its four peers with no direct reciprocal connections, resulting in the arrangement shown in Fig. 3.

Only two of the five workspace nodes in the model have outward connections to other cortical columns, and only three such columns are modelled. In reality all the workspace nodes, including W3, W4, and W5, would have outward connections to other cortical columns, like those of W1 and W2. But three cortical columns alone are sufficient to demonstrate the behaviour of interest here, namely a procession of broadcast workspace states wherein successor states are determined by cortical competition. C2 and C3 represent columns that are close enough to be directly mutually inhibiting and share the same workspace node, while C1 represents a column that is distant from C2 and C3. Mechanisms for competition between distant cortical populations are beyond the scope of the present simulation (see Discussion).

3.1 Structure and Connectivity

Fig. 4 depicts areas W2 and C2 in more detail. The model's other workspace nodes and cortical columns have the same structure. Let's consider the workspace nodes first. A workspace node comprises an excitatory pool (W^+) and an inhibitory pool (W^-). The excitatory pool receives afferents from the output layer of a cortical column (C_{out}) via an access area (A2). Recurrent top-down paths from access areas serve to amplify the activity in the output layer of a cortical column. Efferents run from excitatory workspace neurons (W^+) directly to the input layer of a cortical column (C_{in}), as well as to the other workspace nodes to which it is connected (in this case W3 and W5). Additionally, workspace excitatory neurons stimulate the corresponding inhibitory pool (W^-), which sends efferents to the same workspace nodes (W3 and W5).

Now let's consider the structure of the cortical columns which are not part of the workspace (C1 to C3). In addition to its input and output layers (C_{in} and C_{out}), both of which comprise 256 neurons, each such cortical column includes a pool of 320 non-specific excitatory neurons (C^+) and a pool of 192 non-specific inhibitory

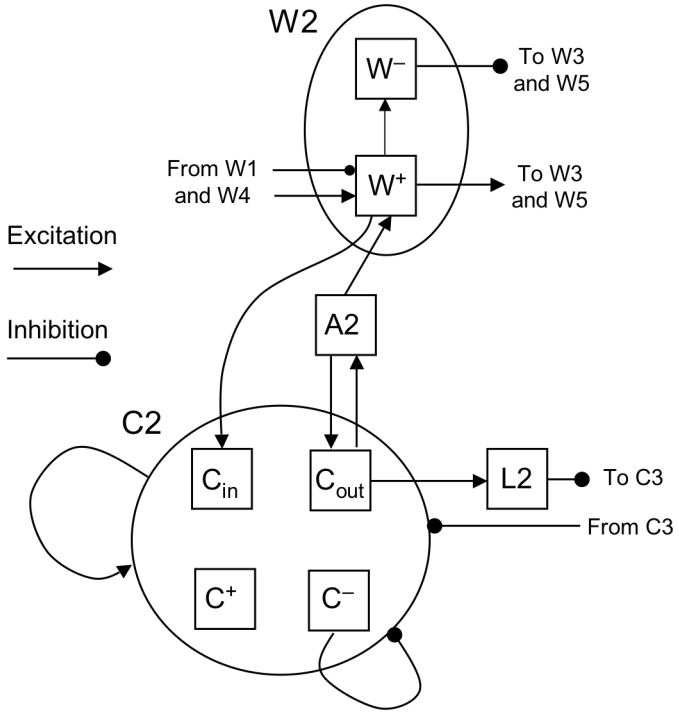

Fig. 4: Detail of areas W2 and C2. C^-, L2, and W^- are inhibitory pools, while the other boxes comprise excitatory neurons. Access area A2 and inhibitory pool L2 are conceptually and anatomically part of column C2.

neurons (C^-). All four sub-areas are recurrently connected to the entire column, with C_{in}, C_{out}, and C^+ making excitatory connections and C^- making inhibitory connections. Additionally, to implement the sort of competitive cortical dynamics that has been used by different authors to model a variety of phenomena (Wang, 2002; Rolls & Deco, 2002; Dehaene, *et al.*, 2003; 2005; Deco & Rolls, 2005), the entire column receives a lateral inhibitory input from its competing neighbour (C3). In the opposite direction, the column stimulates a further pool of 204 inhibitory neurons (L2) which in turn connects to C3. With respect to lateral inhibition, column C3 is a mirror image of C2, while column C1, having no competitors for access to W1, lacks the circuitry for lateral inhibition. Access areas and lateral inhibitory pools are anatomically and functionally included in cortical columns. In Fig. 4, for example, the neurons in C2 are assumed to send out only intra-columnar efferents. A2 and L2 are considered part of the same column as C2, but comprise neurons that project short-range cortico-cortical efferents to nearby columns, in this case C3 and W2.

A critical property of the model is the highly focal nature of the majority of excitatory paths, and in particular of the connections between workspace nodes

and the connections to and from other cortical columns and their access areas (indeed these are all one-to-one in the simulation). This focal character ensures that the spatial properties of a pattern of activation are preserved throughout the workspace as well as in the input and output layers of columns C1 to C3. All inhibitory pathways, by contrast, are diffuse (fully connected with synaptic weights drawn uniformly from [0 1]), as are the recurrent connections within C1 to C3 (also fully connected with initial synaptic weights uniformly drawn from [0 1]). In the overall model, as well as in the individual cortical columns, the proportion of excitatory to inhibitory neurons is approximately four to one, as in real mammalian cortex.

3.2 The 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:

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

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

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

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 $a = 0.02+0.08r$, $b = 0.25-0.05r$, $c = -65$,

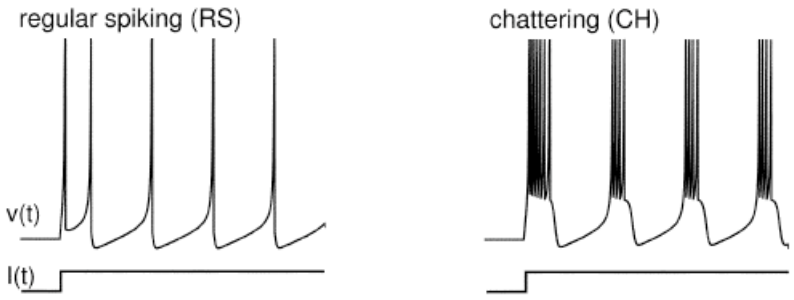

Fig. 5: Varieties of excitatory neurons using Izhikevich's simple model (from Izhikevich, 2003).

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. 5, 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 Φ be the set of all neurons j that fired at time $t–\delta$ where δ is the conductance delay from neuron j to i. Then the input current I for neuron i at time t is given by:

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

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 and conductance delays for the model's various pathways are set out in Table 1. Scaling factors for topographically mapped pathways (eg: to, from, and among workspace areas) are significantly higher than those to and from inhibitory areas and those for recurrent connections within cortical columns to compensate for the correspondingly smaller number of connections per neuron for those pathways.

3.3 Initial Training

The three cortical columns C1, C2, and C3 were subject to an initial period of training while disconnected from the rest of the model, using a variant of spike-timing dependent plasticity (STDP) (Abbott & Nelson, 2000; Song, *et al.*, 2000). STDP is a Hebbian learning rule for spiking neurons that increases the strength of synaptic connections where there is a strong correlation between the timing of

Table 1: Parameters of the Model

F_{WC}	40
F_{AW}	90
F_{WW}	80
F_{CC}	3
F_{CL}	8
F_{LC}	140
F_{WI}	1.2
F_{IW}	1.4
F_{AC}	90
F_{CA}	90
F_{LL}	90

δ_{WC}	10ms
δ_{AW}	20ms
δ_{WW}	5–6ms
δ_{CC}	1–40ms
δ_{CL}	2ms
δ_{LC}	2ms
δ_{WI}	2ms
δ_{IW}	5–6ms
δ_{AC}	2ms
δ_{CA}	2ms
δ_{LL}	2ms

Key to parameters

$F_{\alpha\beta}$ = scaling factor applied to connections from area type α to area type β

$\delta_{\alpha\beta}$ is the conductance delay for connections from area α to β

W = workspace area
I = workspace inhibitory pool
C = cortical column
L = lateral inhibitory pool
A = workspace access area

pre-synaptic spikes and post-synaptic firing, and decreases the strength of connections where this correlation is weak. Details of the (slightly unorthodox) STDP update rule used in the present experiments are relegated to the Appendix, since our only concern here is with its results.

Each column learned to associate the presentation of a certain pattern to its input layer with the later presentation (after 40ms) of a different pattern to its output layer (Gerstner, *et al.*, 1993; Rao & Sejnowski, 2000; Nowotny, *et al.*, 2003; Izhikevich, 2006). Each input and output layer was divided into four separate populations (neuron numbers 1–64, 65–128, 129–192, and 193–256 respectively). The presentation of each input and output pattern involved the excitation, by means of four 10mA pulses at 5ms intervals, of 60% of the neurons in one of those four populations. After training, if a previously seen pattern was presented to a column's input layer, it would respond with the associated pattern in its output layer without requiring further input, after a delay of approximately 40ms (Fig. 6). The input stimuli used to show this consisted of 10ms bursts of random 8mA pulses delivered to the relevant subset of neurons, where the probability of such a neuron receiving a pulse in any given 1ms time step was 0.2. The learned repertoire of associations is the smallest possible – just one per column – that will allow the model to exhibit the desired behaviour, namely a succession of distinct global workspace states. Note that columns C2 and C3 have competing associations for the same input pattern.

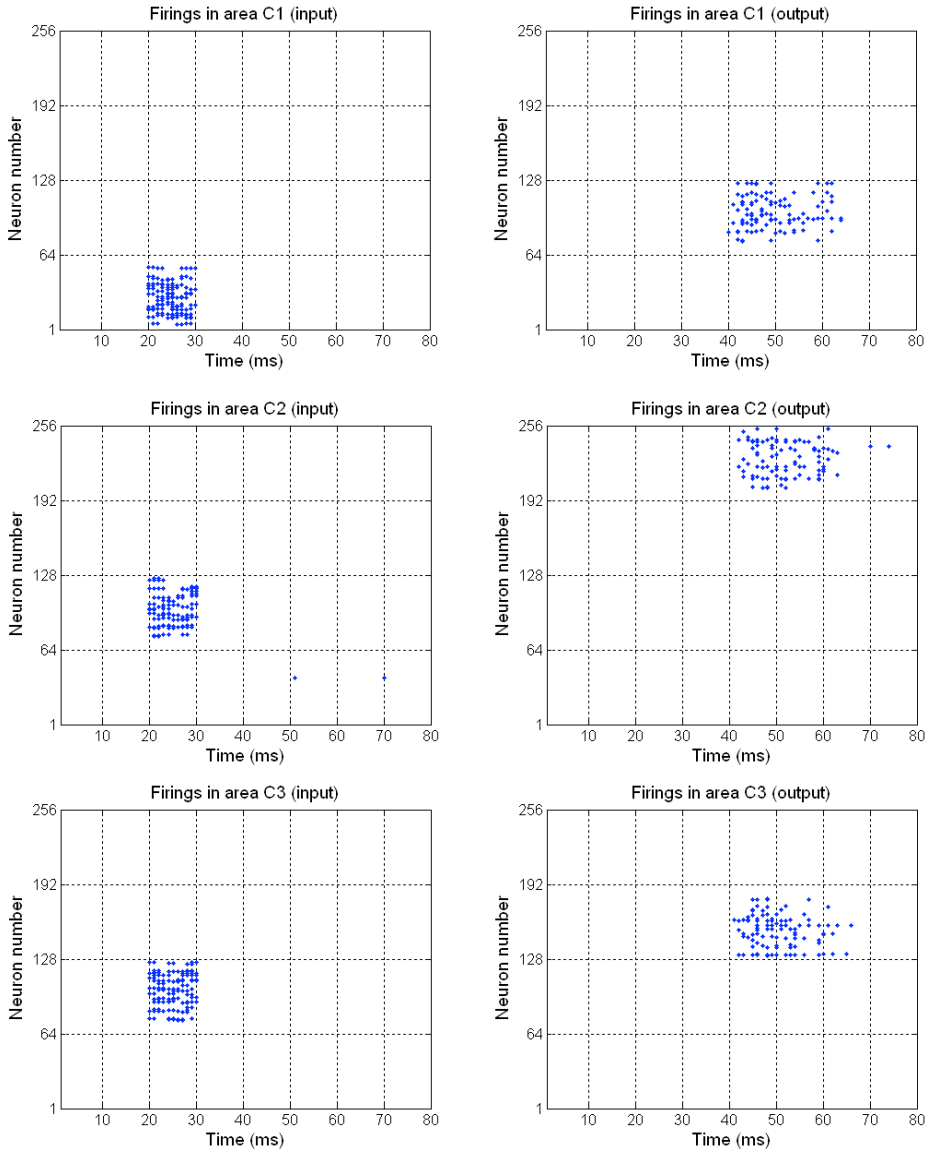

Fig. 6: Cortical associations after initial training. Columns C1 to C3 each store a single input-output pair. The training set is designed so that the output pattern from each column stimulates a unique set of neurons, making it possible to identify the column that caused any given firing in a workspace area.

4 Experimental Results

In each of the experiments described here, an initial stimulus was delivered after 20ms directly to workspace area W1. This took the form of a single set of strong pulses (25mA) to 60% of the sub-population of neurons numbered from 1 to 64. Fig. 7 shows the evolution of areas W1 and W2 during one representative type of trial, while Fig. 8 shows the corresponding evolution of the three cortical columns

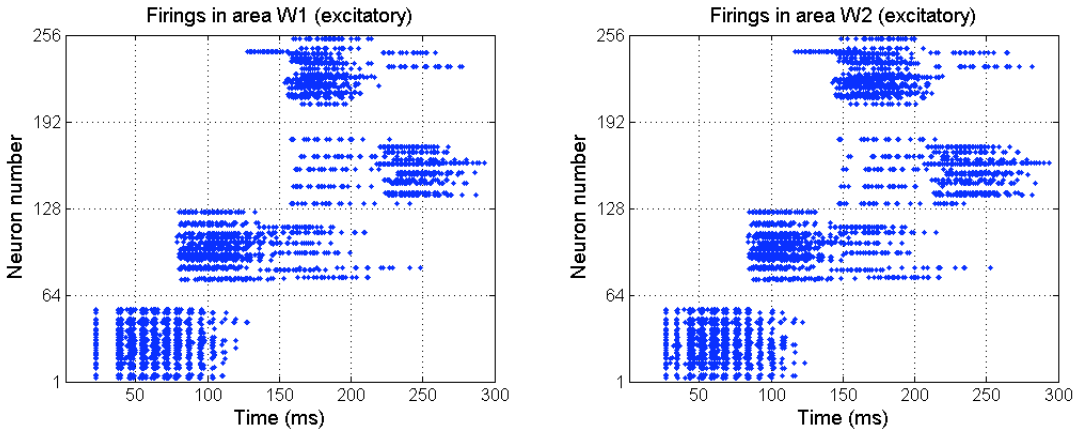

Fig. 7: A sequence of workspace states in a selected trial. Two representative workspace nodes are shown. The pattern of firing is similar in all five. The workspace exhibits a sequence of clearly demarcated states.

C1, C2, and C3 during the same trial. Areas W3 to W5 are not shown, but exhibit the same characteristic pattern as W1 and W2 with a phase difference of up to 5ms. The chain of events depicted in the figures is as follows.

4.1 A Single Trial in Detail

The initial stimulus delivered to W1 at 20ms is transmitted to W2 and W4, thanks to the cortico-cortical connections shown in Fig. 3, from where it is in turn propagated to W3 and W5. Within 20ms this spatial pattern of activation has spread around the ring of workspace nodes, and thanks to the existence of multiple feedback pathways, has set up a self-sustaining reverberation that lasts for approximately 80ms. At the same time, this pattern of activity is transmitted, via workspace areas W1 and W2, to the input layers of cortical columns C1 to C3 (see Figs. 3 & 4).

As shown in Fig. 8 (top right), area C1 begins to show a response to this pattern of activity at around 60ms. Neither of the other cortical columns has an association with strong activation in neurons 1–64, so their output layers remain quiescent. The response from C1 is a burst of activity in neurons 65–128, a pattern which is duly transmitted to workspace area W1. This new pattern of activation now begins to invade the whole workspace, propagating to each of the five workspace areas, and setting up a new self-sustaining reverberation that lasts from around 80ms to 140ms, with some neurons in the population maintaining the reverberation much longer. With the arrival of this new pattern, the old pattern fades thanks to a wave of inhibition (not shown) that spreads around the

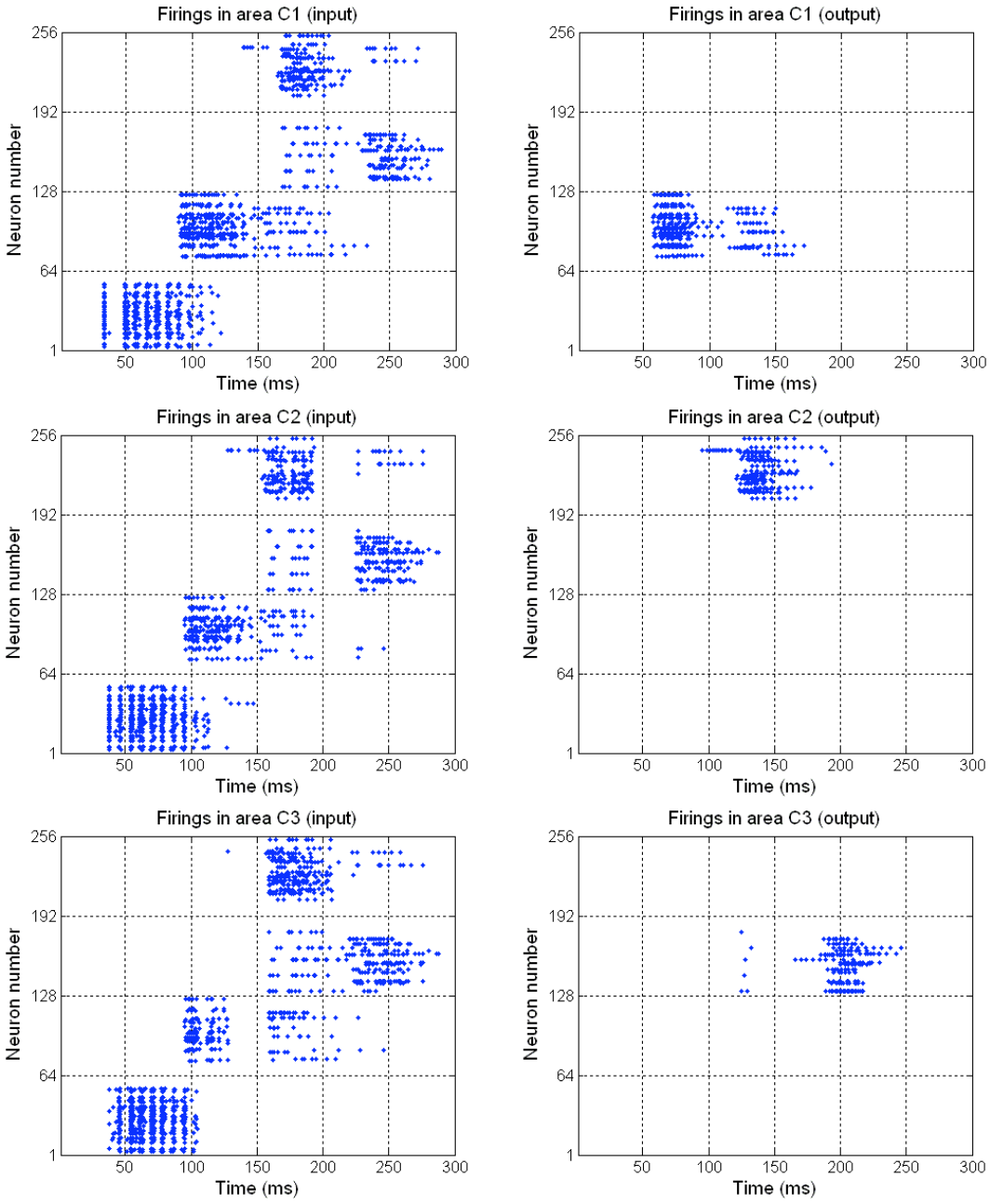

Fig. 8: Cortical column behaviour in a selected trial. The column's input areas (left) echo the state of the workspace. The output areas of columns C1, C2, and C3, in that order, deliver a series of patterns of activation into the workspace. C2 wins the initial competition with C3 (at around 120ms). But C3 gets a turn in the end, thanks to lingering firing in neurons 65–128.

workspace in advance of the new pattern, and by 140ms it has disappeared altogether.

Both C2 and C3 have associations with strong activation in neurons 65–128 (Fig. 8), so a period of competition ensues, owing to their mutually inhibiting

relationship. Though closely matched, and in receipt of near identical input patterns, small statistical differences in the training of columns and C2 and C3 result in an outright winner, namely C2. A strong early response by a subset of neurons in C2's output layer (and by one neuron in particular) is sufficient to excite the inhibitory pool connected to C2 (L2 in Fig. 4), which inhibits C3 and blocks its activity for approximately 50ms, an effect that is reinforced by the top-down amplification of the activity in C2's output layer from area A2. As a consequence, C2 gains exclusive access to the workspace. As Fig. 7 shows, the resulting pattern of strong activation in neurons 193–256 occupies the workspace from approximately 150ms to 210ms.

Despite C2's initial victory, it doesn't have the last word. This is because a subset of the neurons in the workspace, thanks to statistical variations in their parameters, tend to exhibit "chattering" behaviour after initial stimulation by an appropriate set of spikes. These neurons will continue to fire after the rest have become quiescent. This effect is visible throughout the model, as the figures show. (It is noteworthy, however, that the synchronous oscillations induced by the initial stimulus in neurons 1–64 do not seem to give rise to this effect. The reason for this is unclear.)

Of particular interest here is the lingering activity throughout the workspace in neurons 65–128 after C2's initial victory. Because of this, the output layer of C3 exhibits a renewed burst of firing at around 190ms, when the inhibitory influence of C2 has faded (Fig. 8, bottom right). Moreover, by 210ms the reverberating pattern initiated by C2 in workspace neurons 193–256 is also starting to fade, a process accelerated by the transmission of the renewed activation in neurons 129–192 of C3's output layer to workspace area W2, from where it duly propagates throughout the workspace, pushing a wave of inhibition before it. By 290ms this pattern too has faded, and because no further relevant associations are stored in C1 to C3 the whole interconnected system of neuronal populations becomes quiescent.

4.2 Variation Across Trials

A series of 36 trials like the one above was conducted, comprising three trials for each of 12 different training runs. In each training run, the input-output pairs were the same as for the example depicted in Fig. 6. But there were three sources of variation across trainings. First, there was statistical variation in the parameters a,

Table 2: The post-competitive epoch across trials 1 to 18

Trial	Training	Influence of C2 (firings)	Influence of C3 (firings)	Difference (% of total firings)	Winner
1	1	590	127	64.6	C2
2	1	571	161	56.0	C2
3	1	572	17	94.2	C2
4	2	757	364	35.1	C2
5	2	274	718	-44.8	C3
6	2	537	622	-7.3	N/A
7	3	593	0	100.0	C2
8	3	0	542	-100.0	C3
9	3	632	0	100.0	C2
10	4	679	186	57.0	C2
11	4	672	668	0.3	N/A
12	4	764	184	61.2	C2
13	5	579	743	-12.4	N/A
14	5	556	736	-13.9	N/A
15	5	508	731	-18.0	N/A
16	6	126	680	-68.7	C3
17	6	480	590	-10.3	N/A
18	6	532	218	41.9	C2

b, *c*, and *d* used to model individual neurons. Second, initial synaptic strengths prior to the application of STDP were drawn from a uniform distribution over [0 1]. Third, inter-neuronal conductance delays were drawn from a uniform distribution over [1 40]. Additionally, at each 1ms time step throughout training every neuron was assigned a base input current drawn from a Gaussian distribution (mean 0, standard deviation 1). Variation across trials with the same training was the result of the addition of Gaussian noise to the 2mA base input current for each neuron (mean 0, standard deviation 1).

In each trial, the workspace exhibited a succession of well demarcated, stable states, which can be thought of as a series of distinct epochs lasting some 50-60ms. During the first epoch of each trial, the workspace was dominated by the initial stimulus (activity in neurons 1–64), and during the second epoch it was

Table 3: The post-competitive epoch across trials 19 to 36

Trial	Training	Influence of C2 (firings)	Influence of C3 (firings)	Difference (% of total firings)	Winner
19	7	679	658	1.6	N/A
20	7	743	311	41.0	C2
21	7	740	513	18.1	N/A
22	8	723	576	11.3	N/A
23	8	790	503	22.2	N/A
24	8	753	477	22.4	N/A
25	9	750	0	100.0	C2
26	9	820	157	67.9	C2
27	9	671	133	66.9	C2
28	10	457	813	-28.0	C3
29	10	391	778	-33.1	C3
30	10	389	838	-36.6	C3
31	11	184	630	-54.8	C3
32	11	204	524	-44.0	C3
33	11	145	596	-60.9	C3
34	12	196	708	-56.6	C3
35	12	199	708	-56.1	C3
36	12	143	615	-62.3	C3

characterised by activity in neurons 65–128, thanks to the influence of C1. But the character of the third epoch depended on the outcome of the competition between C2 and C3, yielding several qualitatively different types of behaviour. These are summarised in Tables 2 and 3. It should be noted that, thanks to the character of the initial training, each cortical column excites a unique subset of workspace neurons (Fig. 6). Therefore we can clearly identify which cortical column was the cause of any given firing in a workspace area.

The third column of the tables shows the number of C2-influenced firings (neurons 193–256) in the representative workspace area W1 between 155ms and 205ms, a period that more-or-less coincided with the post-competitive epoch in each trial. The fourth column presents the same statistic, but for C3-influenced firings (neurons 129–192). The fifth column shows the difference between C2- and C3-influenced firings as a percentage of the total number of C2- and C3-

influenced firings. If this figure is positive then the competition was won by C2 but if it is negative the winner was C3. The absolute value of this figure indicates the margin of the win. The sixth column shows the winner of the competition. No overall winner is indicated if the margin was less than 25%.

As the tables show, the competition was effectively won by one or other of the cortical columns in 25 out of the 36 trials. C2 is the clear winner in 13 cases, with C3 the clear winner in 12. In the remaining 11 trials both columns gained roughly equal access to the workspace during the period in question. Of especial interest are trainings 2, 3, and 6. For each of these trainings different winners were produced by different trials, although there were no changes in synaptic weights from trial to trial. This suggests a chaos-like sensitivity to small differences at the onset of the competition. The results with training 3 are particularly dramatic. In each of the trials with this training, one or other of the columns managed to silence its rival completely, permitting it no influence on the workspace at all during the period in question (Fig. 9).

5 Discussion

The model described contains three already well-established types of circuit – for sequence learning and retrieval, for competition through mutual inhibition, and for maintaining activation patterns through reverberation. The experimental results presented show that the right combination of this circuitry implements a global neuronal workspace whose behaviour conforms to the four principles set out in Section 2. It sustains and disseminates a spatial pattern, and is sensitive to new patterns that have become established through competitive interaction among cortical populations. As a result, it is capable of exhibiting a succession of distinct, well demarcated stable states. The overall schematic of the model maps cleanly onto the global workspace architecture, making it possible to impose the conscious / non-conscious distinction proposed by global workspace theory onto its information flow. Specifically, the computations carried out locally within the cortical columns C1 to C3 model non-conscious information processing, while patterns of activation that spread throughout the workspace nodes W1 to W5 and into the input layers of the cortical columns model consciously processed information.

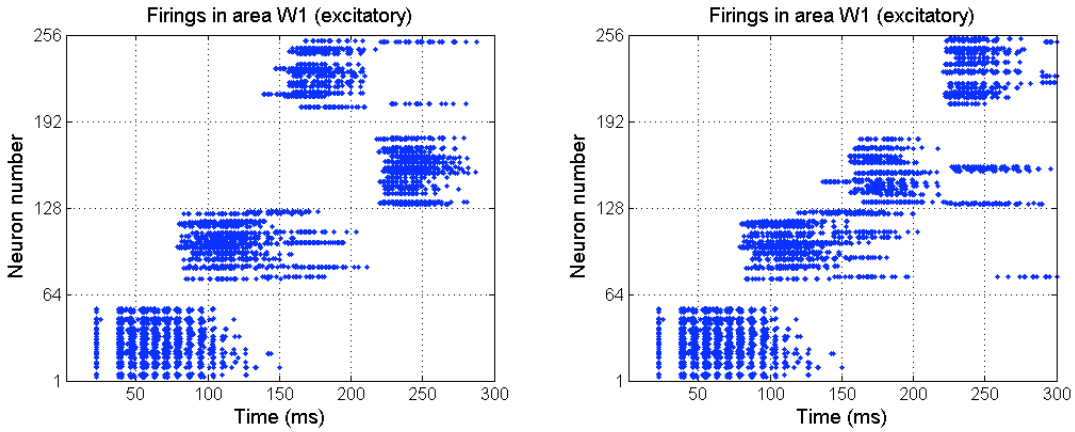

Fig. 9: Trial 7 (left) and Trial 8 (right) yield different sequences of workspace states with the same training and initial stimulus, showing that small statistical variations in intial conditions can result in large, qualitative differences over time.

5.1 An Internal Sensorimotor Loop

In (Shanahan, 2006), a cognitive architecture is proposed in which a global workspace is combined with an internally closed sensorimotor loop. The proposal is in support of the hypothesis that organisms whose brains are endowed with such a loop are capable of rehearsing the consequences of potential actions prior to actually carrying them out (Cotterill, 1998; Hesslow, 2002). The paper presents a computer implementation of the architecture that performs a simple form of cognitively mediated action selection. However, the implementation presented in (Shanahan, 2006), though useful as a proof-of-concept, lacks neurological plausibility, both at the level of the neuron model used and in its employment of a single attractor network to model the global workspace.

The present simulation can be regarded as a neurologically more plausible sequel to the work reported in the earlier paper, and the circuitry described here has the potential to fulfil the same function as the "core circuit" described there. Specifically, if cortical columns learn to associate a landmark sensorimotor state with one or more of the landmark sensorimotor states that might be expected to follow it, the procession of workspace contents can rehearse a trajectory through the organism's sensorimotor space. As a result of this rehearsal, both desirable and undesirable outcomes can be anticipated, and the organism's mechanism for action selection influenced accordingly.

However, to effect a proper search of sensorimotor space, and therefore to carry out planning, the workspace must be capable of revisiting the same sensorimotor state more than once in order to explore alternative outcomes. It has

been shown by Tani (1996) that, though lacking anything analogous to the stack in a conventional computer with a von Neumann architecture, a neurally-based system can in principle search a space of combinatorial structures by exploiting chaotic dynamics. Trials 7 and 8 (Fig. 9) therefore demonstrate a potentially important property of the present model, namely a sufficient degree of sensitivity to small differences in initial conditions for qualitatively indistinguishable states of the workspace to have non-unique successors.

5.2 Shortcomings and Limitations

Needless to say, the model has many shortcomings and limitations that point to the need for further research. For example, the majority of activity in cortical columns C1 to C3 is in the input and output areas, and closely mirrors the activity in the global workspace. The cortical columns, as modelled, are very simple, and there is little in the way of intermediate activity between their inputs and outputs – just a low level of firing in the pools of non-specific neurons (labelled C^+ and C^- in Fig. 4). A richer model, devised to illustrate a wider range of phenomena, would perform more complex cortical computations, and it is easy to imagine expanding each of C1 to C3 to include several hierarchical stages, all of which would, according to global workspace theory, carry out non-conscious information processing. Another drawback of the present simulation is the extent to which a prior structure has been imposed on the workspace (Fig. 3). It would be satisfying if a future, more sophisticated model could demonstrate that the kind of long-range connectivity between remote cortical populations required to realise a global workspace can arise through self-organisation along the lines described in (Izhikevich, *et al.*, 2004).

A further shortcoming is that the only form of competition in the present model is between nearby cortical columns, and it makes no provision for a competition among spatially separated columns with no direct, short-range inhibitory pathways connecting them (such as C1 and C2). A more global mechanism for cortical selection is required for this. One candidate is the type of basal ganglia loop through cortex hypothesised by Redgrave and his colleagues to be implicated in action selection (Redgrave, *et al.*, 1999). Although their modelling work to date has been confined to motor-cortical selection (Prescott, *et al.*, 2006), the anatomical structures and pathways they have emulated seem to be replicated for much of the cortical sheet (Alexander, *et al.*, 1986; Middleton & Strick, 2002;

Postuma & Dagher, 2006). So it seems plausible that they fulfil a similar selectional role throughout, a possibility that has been explored by computational modellers in the context of working memory (Frank, *et al.*, 2001; O'Reilly & Frank, 2006). The incorporation of a similar gating mechanism for workspace access would enhance the present model.

Acknowledgements

Thanks to Bernie Baars and Stanislas Dehaene. Thanks also to Eugene Izhikevich for making his Matlab code publicly available. Finally, thanks to two anonymous reviewers whose many valuable suggestions have improved the paper greatly.

References

Abbott, L.F. & Nelson, S.B. (2000). Synaptic Plasticity: Taming the Beast. *Nature Neuroscience 3*, 1178–1183.

Alexander, G.E., DeLong, M.R. & Strick, P.L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annual Review of Nuroscience 9*, 357–381.

Amit, D.J. & Brunel, N. (1997). Model of Global Spontaneous Activity and Local Structured Activity During Delay Periods in the Cerebral Cortex. *Cerebral Cortex 7*, 237–252.

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 Science 6* (1), 47–52.

Compte, A., Brunel, M., Goldman-Rakic, P.S. & Wang, X.-J. (2000). Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model. *Cerebral Cortex 10*, 910–923.

Constantinedes, C. & Wang, X.-J. (2004). A Neural Circuit Basis for Spatial Working Memory. *Neuroscientist 10* (6), 553–565.

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

Deco, G. & Rolls, E.T. (2003). Attention and Working Memory: A Dynamical Model of Neuronal Activity in the Prefrontal Cortex. *European Journal of Neuroscience 18*, 2374–2390.

Deco, G. & Rolls, E.T. (2005). Neurodynamics of Biased Competition and Cooperation for Attention: A Model with Spiking Neurons. *Journal of Neurophysiology 94*, 295–313.

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

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

Dehaene, S., Naccache, L., Cohen, L., Bihan, D.L., Mangin, J.F., Poline, J.B. & Riviere, D. (2001). Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming. *Nature Neuroscience 4*, 752–758.

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.

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

Frank, M.J., Loughry, B. & O'Reilly, R. (2001). Interactions Between Frontal Cortex and Basal Gaglia in Working Memory: A Computational Model. *Cognitive, Affective, and Behavioral Neuroscience 1* (2), 137–160.

Gerstner, W., Ritz, R., & van Hemmen, W.L. (1993). Why Spikes? Hebbian Learning and Retrieval of Time-Resolved Excitation Patterns. *Biological Cybernetics 69*, 503–515.

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

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

Izhikevich, E.M. (2006). Polychronisation: Computation with Spikes. *Neural Computation 18*, 245–282.

Izhikevich, E.M. (2007). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. MIT Press.

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

Middleton, F.A. & Strick, P.L. (2002). Basal ganglia 'Projections' to the Prefrontal Cortex of the Primate. *Cerebral Cortex 12* (9), 926–935.

Mountcastle, V.B. (1997). The Columnar Organization of the Neocortex. *Brain 120*, 701–722.

Nowotny, T., Rabinovich, M.I., Abarbanal, H.D.I. (2003). Spatial Representation of Temporal Information Through Spike-Timing-Dependent Plasticity. *Physical Review E 68*, 011908.

O'Reilly, R. & Frank, M.J. (2006). Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Computation 18*, 283–328.

Postuma, R.B. & Dagher, A. (2006). Basal Ganglia Functional Connectivity Based on a Meta-Analysis of 126 Positron Emission Tomography and Functional Magnetic Resonance Imaging Publications. *Cerebral Cortex 16* (10), 1508–1521.

Prescott, T.J., Montes-Gonzalez, F., Gurney, K., Humphries, M.D. & Redgrave, P. (2006). A Robot Model of the Bsal Ganglia: Behavior and Intrinsic Processing. *Neural Networks 19* (1), 31–61.

Rao, R.P.N. & Sejnowski, T.J. (2000). Predictive Sequence Learning in Recurrent Neocortical Circuits. In *Proc. Neural Information Processing Systems 12 (NIPS 99)*, ed. S.A.Soller, T.K.Leen, K.-R. Muller, pp. 164–170.

Redgrave, P., Prescott, T.J. & Gurney, K. (1999). The Basal Ganglia: A Vertebrate Solution to the Selection Problem. *Neuroscience 89* (4), 1009–1023.

Rolls, E.T. & Deco, G. (2002). *Computational Neuroscience of Vision*. Oxford University Press.

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.

Sherman, S.M. & Guillery, R.W. (2002). The Role of Thalamus in the Flow of Information to Cortex. *Philosophical Transactions of the Royal Society B 357*, 1695–1708.

Song, S., Miller, K.D. & Abbott, L.F. (2000). Competitive Hebbian Learning Through Spike-Timing-Dependent Synaptic Plasticity. *Nature Neuroscience 3* (9), 919–926.

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

Tani, J. (1996). Model-Based Learning for Mobile Robot Navigation from the Dynamical Systems Perspective. *IEEE Transactions on Systems, Man, and Cybernetics B 26* (3), 421–436.

Wakana, S., Hangyi, J., Nagae-Poetscher, L.M., van Zijl, P.C.M. & Mori, S. (2004). Fiber Tract-Based Atlas of Human White Matter Anatomy. *Radiology 230* (1), 7787.

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

Wang, X.-J. (2002). Probabilistic Decision Making by Slow Reverberation in Cortical Circuits. *Neuron 36*, 955–968.

Appendix: The STDP Learning Rule

The Hebbian learning rule used for the initial training of the three cortical columns C1 to C3 was a form of spike timing dependent plasticity (STDP). After each 1ms time step, the STDP update rule was applied to every neuron in each column. The update rule works as follows. Consider a neuron i that fires at time t_1. We are interested in spikes that arrive at i within a window of ω milliseconds either side of t_1. Suppose that a spike from some neuron j arrives at neuron i at time t_2 such that $-\omega \le \tau \le \omega$, where $\tau = t_2 - t_1$. Then the synaptic weight S of the connection from neuron j to neuron i is adjusted by an amount ΔS, given by the following equation.

$$\Delta S = \begin{cases} 0.68.(S_{\max} - |S|).(1 - \dfrac{|\tau|}{\omega}) & \text{if } \tau \ge 0 \\ -0.68.(S_{\max} - |S|).(1 - \dfrac{|\tau|}{\omega}) & \text{otherwise} \end{cases} \qquad (5)$$

Note that τ depends on the *arrival* time of an incoming spike rather than the firing time of the neuron that delivered it. When, as in the present simulation, there are variable conductance delays, this is clearly the more realistic option, although it is computationally more burdensome since it requires the simulation to maintain more data for each firing. Moreover, as Izhikevich (2006) shows, the interplay of STDP with variable conductance delays (properly treated) can enhance a network's ability to learn spatiotemporal patterns.

For the reported experiments ω = 10ms and S_{max} = 2, giving the characteristic illustrated in Fig. 10. Each column was subjected to two 200ms periods of training in order to learn a single pairing. In each period, the first pattern was presented to the column's input layer as four sets of 10mA pulses, delivered at 20ms, 25ms, 30ms and 35ms. This was followed by the presentation to the output layer of the second, associated pattern in the form of four sets of 10mA pulses delivered at 80ms, 85ms, 90ms, and 95ms. The result of this training for the chosen patterns and their associations is depicted in Fig. 6.

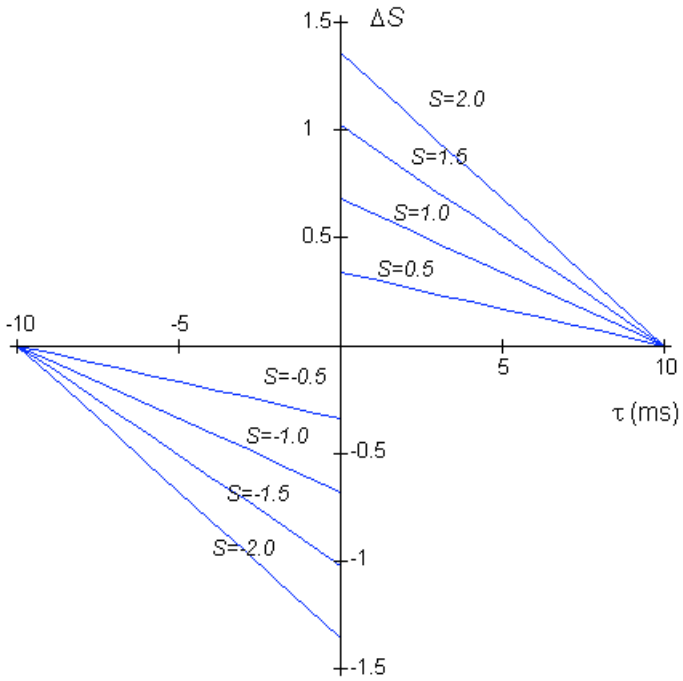

Fig. 10: The characteristic of the STDP update rule. The change in synaptic strength Δ*S* is inversely proportional to the difference τ between the post-synaptic and pre-synaptic spike times, as well as being weighted by the current synaptic strength *S*.

The formulation of the update rule is a little unconventional as it foregoes the usual exponential terms (Song, *et al.*, 2000), instead giving ΔS a linear form. Moreover, the adjustment to S is weighted by S itself, so that synaptic strengths gradually approaches either S_{max} or zero, which does not occur with the more usual form of the rule in which there are sharp cut-offs at S_{max} and zero. The former modification promotes fast training, but also tends to overfit the data very quickly. As such it is well suited to the present application, where only a single pairing of stimuli needs to be learned, as it enables the training process to be completed in just 400ms of simulation time, yet results in a network with realistic statistical and dynamical properties.