Neurocomputational models of working memory

Daniel Durstewitz[1], Jeremy K. Seamans[1] and Terrence J. Sejnowski[1,2]

[1] *Howard Hughes Medical Institute, Salk Institute for Biological Studies, Computational Neurobiology Laboratory, 10010 North Torrey Pines Rd., La Jolla, California 92037, USA*

[2] *Department of Biology, University of California, San Diego, La Jolla, California 92093, USA*

Correspondence should be addressed to T.J.S. (terry@salk.edu)

During working memory tasks, the firing rates of single neurons recorded in behaving monkeys remain elevated without external cues. Modeling studies have explored different mechanisms that could underlie this selective persistent activity, including recurrent excitation within cell assemblies, synfire chains and single-cell bistability. The models show how sustained activity can be stable in the presence of noise and distractors, how different synaptic and voltage-gated conductances contribute to persistent activity, how neuromodulation could influence its robustness, how completely novel items could be maintained, and how continuous attractor states might be achieved. More work is needed to address the full repertoire of neural dynamics observed during working memory tasks.

Working memory is the ability to transiently hold and manipulate goal-related information to guide forthcoming actions[1,2]. The prefrontal cortex (PFC) is the brain structure most closely linked to working memory, based on lesion, local inactivation and brain imaging studies[2–6]. PFC neurons show elevated persistent activity during delayed reaction tasks, when information derived from a briefly presented cue must be held in memory during a delay period to guide a forthcoming response (Fig. 1)[2,5–9]. Persistent delay period activity is often selectively correlated with, and might thus encode, a previously presented cue, a forthcoming response or expected choice situation, or a particular contingency between cue and response[7–13]. If the persistent activity is disrupted either by electrical stimulation or by highly distracting stimuli presented during the delay period, or if it breaks down spontaneously, the animal is highly likely to make an error (Fig. 1b)[2,5,7]. This persistent activity in PFC neurons could carry information about previously encountered stimuli or future responses required to solve working memory tasks. Thus, this type of short-term memory relies on the maintenance of elevated firing rates in specific subpopulations of neurons rather than on synaptic plasticity, which might underlie long-term memory.

The phenomena and mechanisms discussed here are not necessarily unique to the PFC. Sustained, memory-related delay activity is observed in many brain areas, including parietal cortex, inferotemporal cortex, motor areas, hippocampus, and even brain stem[8,14–18]. However, delay activity is more prominent in the PFC than in other areas, and also more robust to interfering stimuli[14], ensuring that delay activity is task-driven and does not passively reflect sensory inputs.

Network models can explain electrophysiological observations and cognitive aspects of working memory tasks. Models have been developed on different levels of abstraction, including highly abstract connectionist models, which neglect the temporal and spatial dynamics of neurons and synapses, firing rate models incorporating some biophysically meaningful time constants, and biophysically detailed models of spiking neurons. Different insights have been gained from each class of models. This review focuses on firing-rate and spiking-neuron models. These models are more closely related to the biophysical mechanisms of neurons and synapses than connectionist models, which have been used to model normal behavior and clinical conditions, learning, neuromodulation and activity profiles in working memory tasks[19–22].

Models of working memory can be broadly classified according to how the persistent activity is generated, although these classes are not mutually exclusive. One mechanism is based on the idea that activity is sustained through strong recurrent excitatory connections in a 'cell assembly'[23]. Most research has been conducted on this class of models. Another hypothesis is that activity circulates in loops, called 'synfire chains'[24], consisting of feedforward-connected subgroups of neurons with no direct feedback links between successive groups. It is also possible that single neurons can maintain activity by membrane currents that allow cellular bistability[25,26]. Another distinction can be made between models that have discrete attractor states representing discrete memory items and models that support continuous attractor states representing continuous variables like space. This review is organized around these distinctions.

Persistent activity through recurrent excitation

Activity may be maintained in a neural network through recurrent excitation. This idea underlies the Hopfield model[27] for storing discrete memory items in the synaptic weight matrix of a network and retrieving them as fixed-point attractors of the activation dynamics. In this model, neurons that collectively encode the same pattern are wired together reciprocally by strong excitatory synaptic weights, forming a cell assembly, whereas neurons that participate in different representations are connected by weak or, in the original Hopfield model, inhibitory synaptic weights (Fig. 2a). These long-term synaptic connection patterns can be acquired through a Hebb-like learning rule that reinforces connections between coactive neurons. A working memory would correspond to the activation of one of these synaptically stored patterns.

Despite its simplicity, the Hopfield model has many proper-

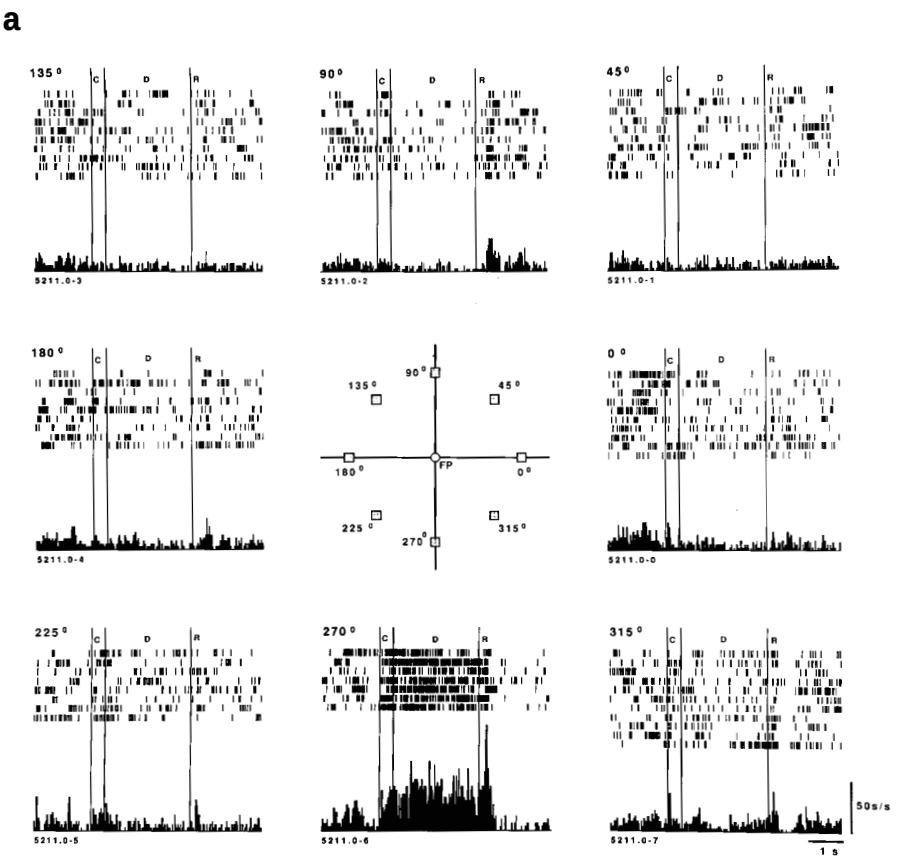

Fig. 1. Delay-period activity recorded in the prefrontal cortex (PFC) *in vivo*. (**a**) Spike frequency histograms showing direction-selective delay activity of a PFC neuron in an oculomotor delayed response task. In this task, a monkey is required to saccade to a briefly presented cue position after a delay, during which the monkey has to fixate a fixation point (FP) in the center of the field. Vertical lines delimit the following task periods: C, cue presentation; D, delay period (no external cue present); R, response period. The neuron shown maintained activity during the delay only if the cue was presented at the bottom (270°) location and was significantly depressed when it was presented in the upper visual field. Reprinted with permission from Fig. 3 in ref. 7. (**b**) Activity of a neuron in three trials of a spatial delayed response task. Bars indicate the period of cue presentation. CR, correct right response; CL, correct left response. During the third trial, a biologically significant distracting stimulus (monkey cries) was presented during the period marked by the dots. In this case, delay activity breaks down, and the monkey fails to respond. If interfering stimuli are biologically less significant, delay period activity can be maintained[14]. Reprinted with permission from Fig. 10 in ref. 5.

ties that are characteristic of human memory, such as similarity-based generalization, fault tolerance and content addressability, which stem from the ability of a recurrent network to retrieve a stored pattern from a degraded or partial input pattern. Subsequent extensions brought this model closer to biology by separating excitatory and inhibitory cell types and enabling the maintenance of activity at physiologically plausible spike rates well below neural saturation levels, with temporal dynamics that can be compared to *in vivo* observations[28–32]. In these models, a neuron is described by two variables: the total synaptic input current, I, and the resulting, monotonically related mean firing rate, R. In Fig. 2b, all the fixed points (self-sustaining states) of one cell assembly of such a network are plotted using variable I (which will be the same for all neurons in a cell assembly if these are identical) versus the afferent input to that assembly, which represents an external stimulus. Suppose the neurons of this cell assembly are stimulated along the afferent input lines (Fig. 2a) with $I_{aff}>0.4$, starting from a silent (subthreshold) network state. The total current will climb until it reaches the firing threshold, at which point recurrent excitation will begin (Fig. 2b and c). Then activity, boosted by recurrent excitation, will rise to the stable suprathreshold fixed point (Fig. 2b, lower broken arrow). When the stimulus is withdrawn ($I_{aff}=0$), the activity will follow the upper suprathreshold curve in Fig. 2b and remain at a suprathreshold level even at $I_{aff}=0$ (a phenomenon called hysteresis). Thus, once driven sufficiently above threshold, network activity will persist in this high state even after removal of the original stimulus (Fig. 2c). Depending on the input, different cell assemblies corresponding to different (and thus selective) persistent states will be stimulated.

The simplicity of this model has made it possible to study in detail the stability conditions for selective states, the number of activity patterns that can be stored, and Hebbian learning with noisy input patterns[28–32]. Moreover, many salient properties of neurons recorded *in vivo* during delay tasks are captured by the model[32–34]. For example, during simulated delayed matching-to-sample tasks, some model neurons respond only during cue presentation, some only during the delay period, and others during both phases; some model neurons respond to intervening stimuli presented during the delay with brief increases in activity rising from the persistent level, others respond with brief decreases, and some model neurons exhibit a 'match enhancement' to the presentation of the target stimulus[32–34]. All these response types are frequently observed *in vivo*[5,7,14].

Empirical basis of cell assembly models

The common assumption underlying the models described above and some biophysically detailed models discussed below is that activity is maintained by recurrent excitation within cell assemblies. Recurrent excitatory connections indeed have been demonstrated anatomically and physiologically *in vitro*[35–37], and the excitatory interactions between cortical neurons during the delay phases of working memory tasks have been probed by simultaneous recordings from multiple neurons[38]. These data do not conclusively show, however, whether recurrent excitation is sufficient to maintain activity.

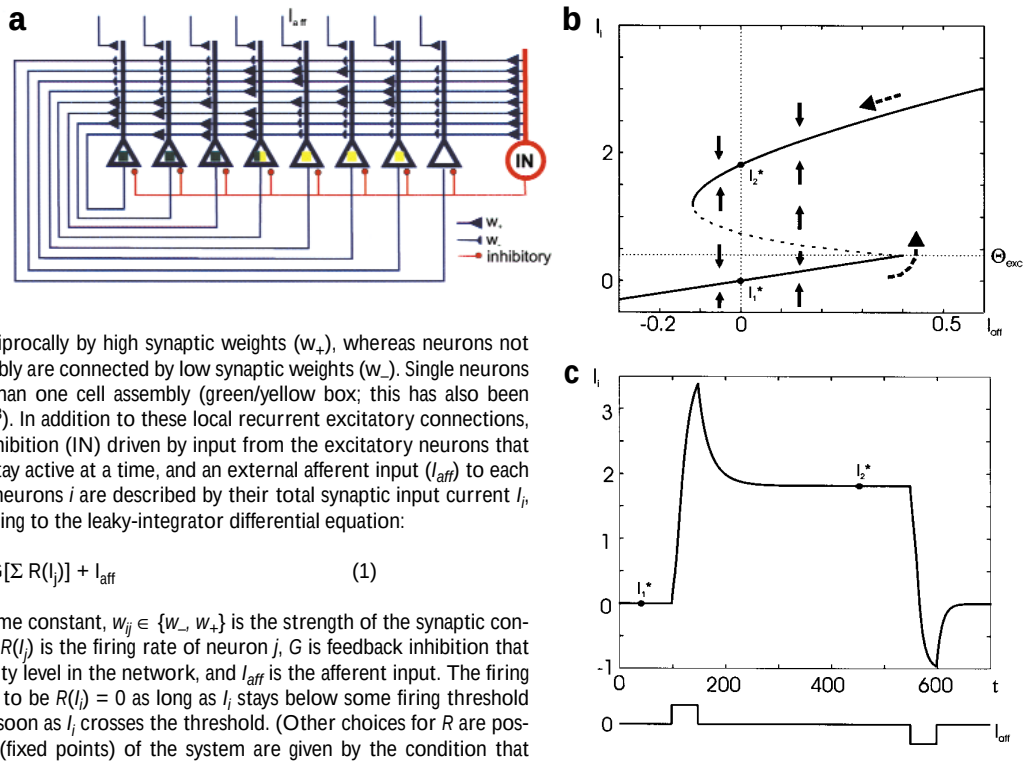

Fig. 2. A firing rate model[28,30,31,34] of delay-period activity in networks of PFC neurons. (**a**) Structure of the network model. Two patterns ('cell assemblies,' green and yellow boxes) coding for two different objects are embedded in the symmetric synaptic weight matrix. Neurons within the same cell assembly are connected reciprocally by high synaptic weights (w_+), whereas neurons not belonging to the same assembly are connected by low synaptic weights (w_-). Single neurons might participate in more than one cell assembly (green/yellow box; this has also been observed experimentally[40,63]). In addition to these local recurrent excitatory connections, there is a global feedback inhibition (IN) driven by input from the excitatory neurons that allows only one pattern to stay active at a time, and an external afferent input (I_{aff}) to each unit in the network. Model neurons i are described by their total synaptic input current I_i, which evolves in time according to the leaky-integrator differential equation:

$$\tau_I \, dI_i/dt = -I_i + \Sigma \, w_{ij} \, R(I_j) - G[\Sigma \, R(I_j)] + I_{aff} \qquad (1)$$

where τ_I is the integration time constant, $w_{ij} \in \{w_-, w_+\}$ is the strength of the synaptic connection from unit j to unit i, $R(I_j)$ is the firing rate of neuron j, G is feedback inhibition that depends on the overall activity level in the network, and I_{aff} is the afferent input. The firing rate of a neuron is assumed to be $R(I_i) = 0$ as long as I_i stays below some firing threshold θ_{exc}, and $R(I_i) = \ln(I_i/\theta_{exc})$ as soon as I_i crosses the threshold. (Other choices for R are possible.) Self-sustaining states (fixed points) of the system are given by the condition that $dI_i/dt = 0$ for all units i, so that the network activity stays constant in time. (**b**) The fixed points (self-sustaining states) of one cell assembly are plotted versus the afferent input (called a bifurcation diagram), assuming that all other neurons in the network are silent. θ_{exc} denotes the firing threshold as defined above. For $I_{aff} < -0.12$, only one stable fixed point exists, which is subthreshold. Similarly, for $I_{aff} > 0.4$, only one stable fixed point exists, but it is suprathreshold. Within the regime $I_{aff} \in [-0.12; 0.4]$, three fixed points coexist; two are stable (solid curves), whereas the third (dashed curve) is unstable. The stability of the fixed points can be determined by evaluating the time derivates (flow field) as given by Eq. 1 for I_i in the vicinity of these points, as indicated by the solid arrows (which here indicate only the direction, not the speed of flow). In the vicinity of the subthreshold and upper suprathreshold fixed points, the flow is toward these points (called their basin of attraction), whereas it is away from the lower suprathreshold fixed point. The subthreshold and upper suprathreshold fixed points are therefore attractors that are stable against perturbations. The two black dots accentuate the two attractor states for $I_{aff} = 0$ (labeled I_1* and I_2*). Different suprathreshold attractor states exist corresponding to the number of different patterns stored in the network. (**c**) Delay-period activity in the network described above. A transient ($\Delta t = 50$) excitatory afferent input ($I_{aff} = 1.0$) at $t = 100$ excites one of the synaptically stored patterns and switches the network into a selective persistent mode, whereas a transient inhibitory afferent input ($I_{aff} = -1.0$) at $t = 450$ terminates persistent activity. (Such inhibitory inputs might originate as a feedback signal from motor systems after goal achievement[77].) Note that total synaptic current of the neurons in one cell assembly is plotted here as a function of time, whereas in (**b**) fixed points of activity are plotted as a function of I_{aff}. The dots mark the same stable steady-state conditions for $I_{aff} = 0$ as in (**b**).

Maintenance of selective activity within a cell assembly also requires that specific synaptic connections be formed that encode the stimulus or response type to be held active in working memory. The specific connections might arise through training and familiarization with the stimuli and response types before testing in delayed-reaction trials. Moreover, some working memory tasks like spatial delayed-response tasks might simply rely on a pre-existing synaptic structure forming topographically organized memory fields in the PFC, even without prior learning[7,39]. However, the need for a pre-existing synaptic structure makes it difficult to explain the ability of humans to retain completely novel stimuli for which no synaptic template might exist yet.

The cell assembly models make a number of specific experimental predictions. For example, as synaptic weights within subgroups of neurons are gradually increased to build up a cell assembly, stable persistent states appear abruptly due to a bifurcation in the network dynamics rather than emerging gradually[28,29]. If persistent states indeed come into existence this way, they should suddenly disappear instead of gradually fading into spontaneous activity as blockade of excitatory synapses with AMPA/NMDA antagonists is gradually increased during successive delay trials *in vivo*. Moreover, if a stimulus is not sufficiently similar to any of the previously learned patterns, activity should decay after removal of the stimulus to the spontaneous level, as may indeed occur in some brain regions or task contexts[40]. On the other hand, overlearned stimulus–response associations seem to evoke less activity in the PFC than relatively novel stimuli[10], indicating that other mechanisms or brain regions might get involved after extensive training.

Low spontaneous and selective high-activity states

In contrast to the model in Fig. 2, PFC neurons *in vivo* are never silent but fire spontaneously at rates of 1–10 Hz between different trials of a working memory task, outside a task context, or even during the delay phases if they are not tuned to the current stimulus or response[5,41–43] (Fig. 1). This raises the question of how spontaneous network activity can remain stable without driving the network automatically into a high-activity state.

This question was addressed using a mean-field approximation to a spiking neuron model where the input from a population of neurons was replaced by its mean and variance[29]. This

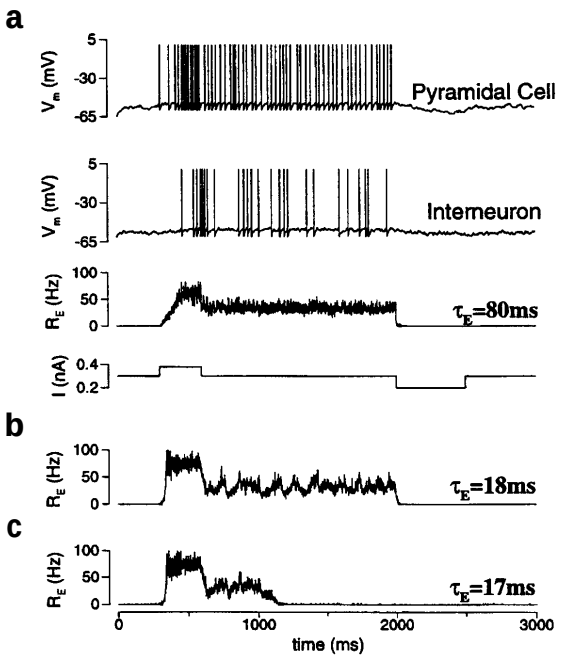

Fig. 3. (Asynchronous) delay activity at physiologically plausible firing rates is not stable if excitatory synapses are too fast. The network model is simulated in the presence of strong recurrent inhibition. The speed of the excitatory synaptic kinetics is varied, whereas the steady-state synaptic drive and the mean firing rate are maintained. (**a**) With a decay time constant for the excitatory synapses of τ_E = 80 ms, a brief current injection turns the network on to a persistent state, with a mean firing rate of the excitatory neurons of R_E ~ 33 Hz. (**b**) With τ_E = *18* ms, the persistent state is still stable, but the firing rate shows large fluctuations in time. (**c**) With τ_E = 17 ms, the fluctuations eventually bring $R_E(t)$ too close to zero, and the network returns to the rest state. Reprinted with permission from ref. 44, Fig. 8.

analysis showed that global low spontaneous activity could be another stable state if there is a slight dominance of local inhibition over excitation, or if the integration time constant of the excitatory neurons is much slower than the inhibitory ones. However, this model assumed that only a small fraction of the neurons in the network participate in encoding a given persistent pattern ('sparse coding'), in contrast to the *in vivo* finding that a large percentage of the recorded neurons are active during any given delay period[5,9,14]. The stability of spontaneous states under these conditions needs to be examined further.

An intuitive way to see how spontaneous activity could be stable is to focus on the dotted I_{aff} = 0 line in Fig. 2b, and to note that transient suprathreshold current fluctuations will not by themselves drive one of the cell assemblies into the upper stable fixed point as long as these presumably random and non-selective fluctuations stay below the curve marked by the unstable fixed point. Such fluctuations might be induced by time-varying external inputs with a subthreshold mean (for example, $<I_{aff}>$ = 0) that occasionally drive neurons across threshold and thus elicit spiking. Hence, spontaneous activity might result from random non-selective perturbations around a subthreshold resting state that do not get reinforced sufficiently by recurrent excitation.

Synaptic basis of persistent activity

Firing rate models provide insight into many aspects of attractor networks but generally ignore the wide range of biophysical time scales in cortical neurons, the consequences of single spikes for dynamics, and the specific contributions of voltage-gated and synaptic ion channels to delay activity. For example, it is not at all clear whether robust delay activity, at frequencies as low as the 15–20 Hz observed *in vivo*[5,7,9,14], can actually be achieved by recurrent excitation within a local network with realistic synaptic time constants[44,45]. Given the fast decay of AMPA receptor currents, only much higher firing frequencies, on the order of the inverse of the AMPA current time course (> 50 Hz), might be reasonably robust to noise or distractors.

Recent biophysical models with spike output based on conductance changes have shown that AMPA currents alone are indeed insufficient to maintain robust delay activity at physiologically realistic rates if slower negative feedback mechanisms like synaptic short-term depression are present[44]. However, NMDA receptor currents, which last over eighty milliseconds, could enable robust delay activity in the 15–40 Hz range by providing a nearly constant synaptic drive[45,46] (Fig. 3). These results emphasize the critical importance of NMDA currents to normal synaptic function, apart from their contribution to synaptic plasticity. In contrast, increasing the relative contribution of AMPA versus NMDA receptor currents leads to more synchronized and less robust delay activity that is more vulnerable to noise and interfering input[44–46]. Interestingly, the PFC has the highest NMDA receptor density of all cortical areas[47], forming a possible basis for its special role in sustaining delay activity. These models predict that partial blockade of NMDA currents should diminish delay activity, consistent with observations that NMDA blockers interfere with working memory and delay activity[48,49]. In contrast, partial blockade of AMPA currents should render delay activity more robust to distractors[14,45].

Neuromodulation of working memory activity

Neuromodulators might alter the processing mode of prefrontal networks to adjust them to specific tasks. The best-studied neuromodulatory system for working memory in the PFC is the dopaminergic input from midbrain neurons in the ventral tegmental area and substantia nigra pars compacta. Dopaminergic activity increases during working memory tasks[50,51] and optimal stimulation of D1 receptors in the PFC is essential for performance involving working memory[52–55], but the function of dopamine in working memory is not yet known.

Dopamine has multiple effects on voltage-gated and synaptic currents in PFC neurons *in vitro*. For example, it enhances persistent Na^+ and NMDA-receptor conductances[56–59]. When these ionic and synaptic effects of dopamine were included in biophysically detailed network models of PFC neurons, the robustness of delay-period activity to distracting stimuli and noise was greatly enhanced[34,45]. This was a consequence of nonlinear interactions between the dopamine-regulated currents and network activity that strengthened the currently active cell assembly but suppressed spontaneous activity and competing, currently inactive assemblies. By increasing the robustness of working memory representations, dopamine might ensure that actions remain directed toward behaviorally relevant goals over extended periods of time in the face of competing and distracting stimuli.

These models predict that local application of dopamine agonists at low doses in the PFC should make delay-type activity more robust to distractor stimuli, whereas high doses might overstabilize representations so that they persist even between trials, result-

ing in response perseveration—the inability of an animal to switch to a new response type[34,45]. Another prediction is related to the observation that a dopamine-induced increase in $GABA_A$ receptor currents was necessary to suppress the spontaneous activation of task-irrelevant representations in the model. As a consequence, partial blockade of $GABA_A$ receptor currents *in vivo* in the PFC might enhance delay activity but should reduce working memory performance[45], in line with experimental results[60].

Recent evidence suggests that the effects of dopamine on prefrontal neurons might depend on time, agonist concentration and the receptor subtype activated[57–59,61], possibly producing complex outcomes that should be addressed in future modeling studies.

Persistent activity through synfire chains

An alternative way to sustain activity locally at delay-period rates as observed *in vivo* is a 'synfire chain'—a wave of synchronous activity that travels through feedforward-connected subgroups of neurons arranged in a chain and which might be maintained through closed loops within the chain[24,62]. Thus, activity is propagated from subpopulation to subpopulation through asymmetric connections as opposed to the dense reciprocal connectivity that maintains activity within cell assemblies.

Activity in synfire chains might seem vulnerable to either decay or blowup, but stable, noise-tolerant propagation of activity packets through subpopulations of neurons is possible as long as the total activity and dispersion in spiking times stay within some reasonable limits forming a basin of attraction[62]. During delay periods synchronous spiking occurs in the motor cortex with higher than chance frequency and different subgroups of neurons become synchronized during different task periods[63], consistent with activity in a synfire chain.

These experimental findings do not, however, rule out any of the competing models. Moreover, synfire chains might require more neurons to sustain activity because the lack of a dense recurrent connectivity makes sustained activity along the chain more sensitive to loss of feedforward connections. Activity in a synfire chain might also be less robust to interference than NMDA-mediated recurrent excitation since it requires some synchrony in spiking times[62] which makes it more vulnerable to noise and GABAergic inhibition[44,46].

Working memory through cellular bistability

Models based on recurrent excitation and synfire chains can deal with novel stimuli or combinations of features only through synaptic learning, which might be too slow to account for our ability to generalize immediately to novel stimuli in trial-unique working-memory tasks[64]. A possible solution to this problem is based on cellular bistability, in which single neurons have two different stable states, one resting state and one continuously spiking 'up' state. Thus, this mechanism does not rely on a specific synaptic matrix formed by prior learning to maintain activity. The up state could either be sustained completely independently of synaptic inputs, by voltage/Ca^{2+}-gated membrane currents, or might exist only in the presence of sufficient synaptic drive. Input-independent cellular bistability has been used in connectionist and spiking neuron models to store novel information and represent memory items consisting of arbitrary combinations of features[20,65,66]. One disadvantage of input-independent cellular bistability is that the state of the single neurons maintaining the representation may be more sensitive to distractors or noise than if delay activity depended also on synaptic feedback from other neurons.

The nonlinear current–voltage relationship of the NMDA receptor could provide a basis for selective bistability that does not depend on a specific pattern of synaptic connectivity, although it does depend on synaptic inputs[26]. Because of the voltage dependence of the NMDA receptor conductance, the current is zero at its reversal potential around 0 mV, approaches zero again for very negative membrane potentials, and peaks in between. This property, in the right combination with AMPA and $GABA_A$ receptor currents, can produce two stable fixed points of the membrane potential (Fig. 4a), where the higher fixed point corresponds to a form of NMDA-receptor-mediated dendritic plateau potential (Fig. 4b). An arbitrary subgroup of neurons could be locked into the higher fixed point, such that selective persistent activity is due to recurrent synaptic drive and to the dendrites of this subgroup settling at a higher membrane voltage, which removes the Mg^{2+} block of dendritic NMDA receptor channels. Neurons in the network not depolarized to this level would remain at the lower fixed point of the membrane potential and hence have much less NMDA receptor

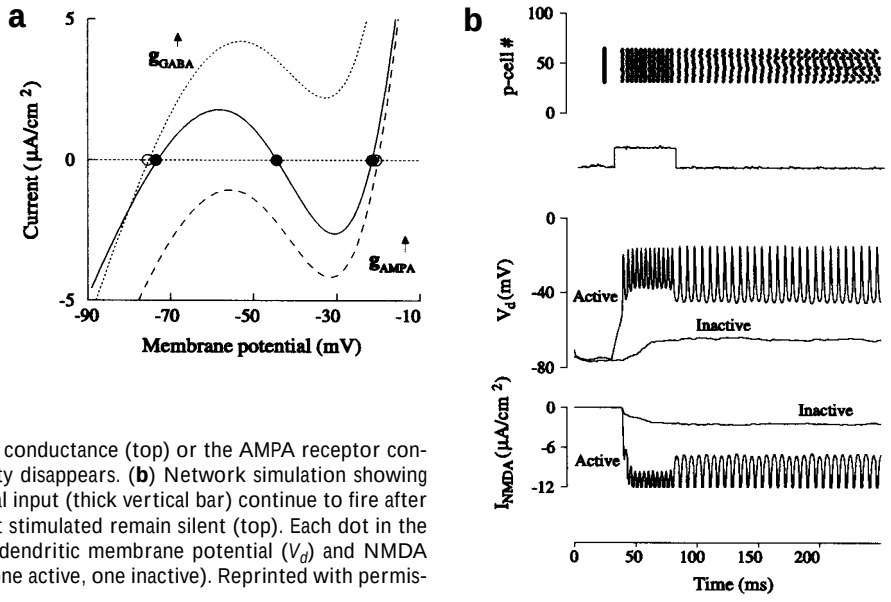

Fig. 4. Maintenance of selective working memory by NMDA-receptor-induced cellular bistability. The structure of the network was the same as in Fig. 2a, with the important difference that all recurrent synaptic weights were the same; that is, there was no prestructured synaptic matrix. (**a**) Steady-state current–voltage curve for the neuron's synaptic conductances, $I = g_{GABA}(V - V_{GABA}) + g_{AMPA}V + g_{NMDA}\,V/(1 + 0.15\,e^{-0.08V})$. Solid dots mark three zero crossings (fixed points) in the solid middle curve; two of these occur at voltages where the slope is positive and the neuron is therefore bistable. If the GABA receptor conductance (top) or the AMPA receptor conductance (bottom) is increased, bistability disappears. (**b**) Network simulation showing that pyramidal cells that received external input (thick vertical bar) continue to fire after input ceases, whereas cells that were not stimulated remain silent (top). Each dot in the raster plot represents a spike. Bottom, dendritic membrane potential (V_d) and NMDA current (I_{NMDA}) for two pyramidal cells (one active, one inactive). Reprinted with permission from ref. 26, Fig. 1.

current, while receiving the same synaptic input.

There is some experimental evidence for cellular bistability in prefrontal neurons *in vivo* mediated by NMDA receptors and modulated by dopamine (H. Moore *et al.*, *Soc. Neurosci. Abstr.* 24, 823.8, 1998; B.L. Lewis & P. O'Donnell, *Soc. Neurosci. Abstr.* 25, 664.3, 1999). Single PFC neurons *in vitro* also show bistability following application of muscarinic agonists that enhance an afterdepolarizing, Ca^{2+}-activated mixed cationic current[67]. This current can sustain spiking that outlives the short period of stimulation. Thus, acetylcholine, a neuromodulator that is also involved in working memory[68], might promote cellular bistability in PFC neurons that is independent from synaptic input. However, more experimental evidence is needed to support the idea of cellular bistability and its possible role in persistent activity in the PFC, on its possible ionic basis, and its dependence on neuromodulators and synaptic inputs. For example, whether NMDA receptor currents can indeed produce a dendritic plateau potential[26] could be tested *in vitro*.

Models with continuous attractors

The recurrent networks discussed so far have connectivity that supports discrete, multistable attractor states, which are points in the space spanned by the firing rates of the neurons and/or their spatial positions in the network. To address the question of how continuously-valued variables like spatial position or stimulus frequency can be maintained in memory, recurrent network models have been developed that approximate a continuum of stable states that have the topology of lines and surfaces in time-averaged quantities like the firing rates. That is, activity states of these systems are stable against perturbations that push the system away from these lines or surfaces, but not necessarily to perturbations that move the system along these lines or surfaces. Wilson and Cowan[69] and Amari[70] were among the first to explore such dynamics in firing rate models.

An example of a system with continuous attractors is the compass cells found in various limbic areas of the rat brain, which represent the direction of the rat's head. This activity persists in the dark and does not depend on visual input[71]. A firing rate network model with activity states that can move around a circle shares several properties of compass cells[72]. However, without visual input, there is a slow drift in the spatial position of the activity profile, which is reproduced in the model when some noise is added, indicating that the state of the system is indeed not stable against perturbations along the attractor continuum. Another example is the spatial position of a cue in the oculomotor delayed response task[7] (Fig. 1a), which could be encoded by a continuum of spatially tuned activity profiles in the PFC[73]. This latter model also used cellular bistability, ensuring stability against perturbations along the spatial quasi-continuum, although the bistability *per se* was not needed to maintain activity. In all these models, the continuity of attractor states is achieved by synaptic connections that are symmetrical along points in space and a decreasing function of the spatial distance between neurons, such that activity profiles can be localized anywhere in the neural space.

A spatial continuum of activity patterns constitutes one way to represent continuous variables like location or direction. Another possibility is to represent continuous-valued sensory or motor attributes by a continuum of persistent firing rates. For example, PFC neurons recorded *in vivo* in a parametric working memory task encode the flutter frequency of tactile inputs monotonically with firing rate during the delay[74]. Similar activity patterns occur in integrator neurons in the oculomotor system, which linearly encode in their firing rates the current eye position even without persistent visual input[16]. This behavior was recently reproduced in a biophysically detailed model that has a continuous range of quasi-stable states of persistent firing[75]. A nearly continuous range of self-sustaining activity levels was achieved by precisely tuning all synapses such that, with rising activity levels, the increasing saturation in recurrent synaptic inputs was compensated by recruiting more and more neurons into the active state, resulting in a fine-tuned balance. However, this model is highly sensitive to the exact tuning of the synapses, and there may be more robust ways to represent a continuum of persistent firing rates.

Conclusions

Network models have provided general insights into the specific mathematical conditions that allow networks to have multiple selective and stable persistent memory states in addition to non-selective spontaneous states. Detailed realistic models were used recently to explore the specific ionic mechanisms that may underlie robust persistent activity and working memory performance. By making much closer contact to *in vitro* and *in vivo* data than earlier models, it has also been possible to make physiologically more specific predictions. Moreover, new ideas have been introduced for fast and flexible coding in working memory, for how neuromodulators affect working memory, and for continuous attractors that represent continuous variables.

The different cellular and network mechanisms reviewed here are not mutually exclusive, but may co-occur in the PFC and other brain structures to allow a large variety of strategies for flexible coding and manipulation of information. For example, cellular bistability could be used to actively maintain novel items, but might not be sufficiently robust to distractors and noise. Through mechanisms for synaptic plasticity, more permanent representations of these items might be formed that enhance robustness of sustained activity and enable fast processing at lower, metabolically economical firing rates.

Holding onto information, and the stimulus-selective, sustained activity associated with it, is just one aspect of working memory. During *in vivo* recordings, transitions between different types of activity are observed, with stimulus-related activity often decreasing and response- or expectancy-related activity slowly increasing during the delay[8,10,43,74]. These findings reveal a much richer dynamic repertoire than has been addressed so far with models. It is also not clear how these electrophysiological phenomena relate to cognitive processes in the PFC. Another open issue is how performance in delayed-reaction tasks can be acquired through a series of conditioning procedures as used in animal experiments[10]. This question has been addressed in connectionist models[20] but not yet in biophysically realistic models. Finally, neural models should be extended to provide insights into the dynamics underlying higher cognitive functions of the PFC that are based on working memory, such as planning and problem solving[4,76], thus fully addressing the question of 'working with memory' in the context of goal-directed behavior.

Acknowledgements

D.D. was funded through a research stipend from the Deutsche Forschungsgemeinschaft (DU 354/1-1). J.K.S and T.J.S. were supported by the Howard Hughes Medical Institute. Thanks to Emilio Salinas, Paul Tiesinga, Sabine Windmann and Kechen Zhang for comments on the manuscript.

Received 9 June; accepted 2 October 2000

1. Baddeley, A. *Human Memory* (Lawrence Erlbaum, Hove, UK, 1990).
2. Fuster, J. M. *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe* 3rd edn. (Lippincott-Raven, New York, 1997).
3. Cohen J. D. *et al.* Temporal dynamics of brain activation during a working memory task. *Nature* 386, 604–608 (1997).
4. Dehaene, S., Jonides, J., Smith, E. E. & Spitzer, M. in *Fundamental Neuroscience* (eds. Zigmond, M. J., Bloom, F. E., Landis, S. C., Roberts, J. L. & Squire, L. R.) 1543–1564 (Academic, San Diego, California, 1999).
5. Fuster, J. M. Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *J. Neurophysiol.* 36, 61–78 (1973).
6. Goldman-Rakic, P.S. in *Models of Information Processing in the Basal Ganglia* (eds. Houk, J. C., Davis, J. L. & Beiser, D. G.) 131–148 (MIT Press, Cambridge, Massachusetts, 1995).
7. Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61, 331–349 (1989).
8. Quintana, J., & Fuster, J. M. From perception to action: temporal integrative functions of prefrontal and parietal neurons. *Cereb. Cortex* 9, 213–221 (1999).
9. Rainer, G., Asaad, W. F. & Miller, E. K. Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* 393, 577–579 (1998).
10. Asaad, W. F., Rainer, G. & Miller, E. K. Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 21, 1399–1407 (1998).
11. Boussaoud, D. & Wise, S. P. Primate frontal cortex: effects of stimulus and movement. *Exp. Brain Res.* 95, 28–40 (1993).
12. Fuster, J. M., Bodner, M. & Kroger, J. K. Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* 405, 347–351 (2000).
13. Rao, S. C., Rainer, G. & Miller, E. K. Integration of what and where in the primate prefrontal cortex. *Science* 276, 821–824 (1997).
14. Miller, E. K., Erickson, C. A. & Desimone, R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* 16, 5154–5167 (1996).
15. Constantinidis, C. & Steinmetz, M. A. Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *J. Neurophysiol.* 76, 1352–1355 (1996).
16. McFarland, J. L. & Fuchs, A. F. Discharge patterns in nucleus prepositus hypoglossi and adjacent medial vestibular nucleus during horizontal eye movement in behaving macaques. *J. Neurophysiol.* 68, 319–332 (1992).
17. Miller, E. K., Li, L. & Desimone, R. Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J. Neurosci.* 13, 1460–1478 (1993).
18. Watanabe, T. & Niki, H. Hippocampal unit activity and delayed response in the monkey. *Brain Res.* 325, 241–254 (1985).
19. Braver, T. S., Barch, D. M. & Cohen, J. D. Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. *Biol. Psychiatry* 46, 312–328 (1999).
20. Guigon, E., Dorizzi, B., Burnod, Y. & Schultz, W. Neural correlates of learning in the prefrontal cortex of the monkey: a predictive model. *Cereb. Cortex* 5, 135–147 (1995).
21. Moody, S. L., Wise, S. P., Di Pellegrino, G. & Zipser, D. A model that accounts for activity in primate frontal cortex during a delayed matching-to-sample task. *J. Neurosci.* 18, 399–410 (1998).
22. Zipser, D., Kehoe, B., Littlewort, G. & Fuster, J. A spiking network model of short-term active memory. *J. Neurosci.* 13, 3406–3420 (1993).
23. Hebb, D. O. *The Organization of Behavior* (Wiley, New York, 1949).
24. Abeles, M. *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge Univ. Press, Cambridge, 1991).
25. Marder, E., Abbott, L. F., Turrigiano, G. G., Liu, Z. & Golowasch, J. Memory from the dynamics of intrinsic membrane currents. *Proc. Natl. Acad. Sci. USA* 93, 13481–13486 (1996).
26. Lisman, J. E., Fellous, J. M. & Wang, X.-J. A role for NMDA-receptor channels in working memory. *Nat. Neurosci.* 1, 273–275 (1998).
27. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* 79, 2554–2558 (1982).
28. Amit, D. J. & Brunel, N. Learning internal representations in an attractor neural network with analogue neurons. *Network Comput. Neural Systems* 6, 359–388 (1995).
29. Amit, D. J. & Brunel, N. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex*. 7, 237–252 (1997).
30. Amit, D. J. & Tsodyks, M.V. Quantitative study of attractor neural networks retrieving at low spike rates: I. Substrate-spikes, rates and neuronal gain. *Network* 2, 259–273 (1991).
31. Amit, D. J. & Tsodyks, M.V. Quantitative study of attractor neural networks retrieving at low spike rates: II. Low-rate retrieval in symmetric networks. *Network* 2, 275–294 (1991).
32. Amit, D. J., Brunel, N. & Tsodyks, M. V. Correlations of cortical hebbian reverberations: theory versus experiment. *J. Neurosci.* 14, 6435–6445 (1994).
33. Amit, D. J., Fusi, S. & Yakovlev, V. Paradigmatic working memory (attractor) cell in IT cortex. *Neural Comput.* 9, 1071–1092 (1997).
34. Durstewitz, D., Kelc, M. & Güntürkün, O. A neurocomputational theory of the dopaminergic modulation of working memory functions. *J. Neurosci.* 19, 2807–2822 (1999).
35. Gonzalez-Burgos, G., Barrionuevo, G. & Lewis, D. A. Horizontal synaptic connections in monkey prefrontal cortex: an in vitro electrophysiological study. *Cereb. Cortex* 10, 82–92 (2000).
36. Markram, H., Lübke, J., Frotscher, M., Roth. A. & Sakmann, B. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol. (Lond.)* 500, 409–440 (1997).
37. Melchitzky, D. S., Sesack, S. R., Pucak, M. L. & Lewis, D. A. Synaptic targets of pyramidal neurons providing intrinsic horizontal connections in monkey prefrontal cortex. *J. Comp. Neurol.* 390, 211–224 (1998).
38. Funahashi, S. & Inoue, M. Neuronal interactions related to working memory processes in the primate prefrontal cortex revealed by cross-correlation analysis. *Cereb. Cortex* 10, 535–551 (2000).
39. Goldman-Rakic, P.S. Topography of cognition: parallel distributed networks in primate association cortex. *Annu. Rev. Neurosci.* 11, 137–156 (1988).
40. Miyashita Y. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* 335, 817–820 (1988).
41. Rosenkilde, C. E., Rosvold, H. E. & Mishkin, M. Time discrimination with positional responses after selective prefrontal lesions in monkeys. *Brain Res.* 210, 129–144 (1981).
42. Sawaguchi, T., Matsumura, M. & Kubota, K. Effects of dopamine antagonists on neuronal activity related to a delayed response task in monkey prefrontal cortex. *J. Neurophysiol.* 63, 1401–1412 (1990).
43. Rainer, G., Rao, S. C. & Miller, E. K. Prospective coding for objects in primate prefrontal cortex. *J. Neurosci.* 19, 5493–5505 (1999).
44. Wang, X. J. Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J. Neurosci.* 19, 9587–9603 (1999).
45. Durstewitz, D., Seamans, J. K. & Sejnowski T. J. Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *J. Neurophysiol.* 83, 1733–1750 (2000).
46. Compte, A., Brunel, N., Goldman-Rakic, P. S. & Wang, X.-J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* 10, 910–923 (2000).
47. Scherzer, C. R. *et al.* Expression of N-methyl-D-aspartate receptor subunit mRNAs in the human brain: hippocampus and cortex. *J. Comp. Neurol.* 390, 75–90 (1998).
48. Aura, J. & Riekkinen, P. J. Blockade of NMDA receptors located at the dorsomedial prefrontal cortex impairs spatial working memory in rats. *Neuroreport* 10, 243–248 (1999).
49. Dudkin, K. N., Kruchinin, V. K. & Chueva, I. V. Effect of NMDA on the activity of cortical glutaminergic structures in delayed visual differentiation in monkeys. *Neurosci. Behav. Physiol.* 27, 153–158 (1997).
50. Schultz, W., Apicella, P. & Ljungberg, T. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* 13, 900–913 (1993).
51. Watanabe, M. Reward expectancy in primate prefrontal neurons. *Nature* 382, 629–632 (1996).
52. Müller, U., von Cramon, D. Y. & Pollmann, S. D1- versus D2-receptor modulation of visuospatial working memory in humans. *J. Neurosci.* 18, 2720–2728 (1998).
53. Sawaguchi, T. & Goldman-Rakic, P. S. The role of D1-dopamine receptor in working memory: local injections of dopamine antagonists into the prefrontal cortex of rhesus monkeys performing an oculomotor delayed-response task. *J. Neurophysiol.* 71, 515–528 (1994).
54. Seamans, J. K., Floresco, S. B. & Phillips, A. G. D1 receptor modulation of hippocampal-prefrontal cortical circuits integrating spatial memory with executive functions in the rat. *J. Neurosci.* 18, 1613–1621 (1998).
55. Zahrt, J., Taylor, J. R., Mathew, R. G. & Arnsten, A. F. T. Supranormal stimulation of D_1 dopamine receptors in the rodent prefrontal cortex impairs spatial working memory performance. *J. Neurosci.* 17, 8528–8535 (1997).
56. Seamans, J.K., Durstewitz, D., Christie, B., Stevens, C. F. & Sejnowski, T. J. Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons. *Proc. Natl. Acad. Sci. USA* (in press).
57. Gorelova, N. A. & Yang, C. R. Dopamine D1/D5 receptor activation modulates a persistent sodium current in rat prefrontal cortical neurons in vitro. *J. Neurophysiol.* 84, 75–87 (2000).
58. Yang, C. R. & Seamans, J. K. Dopamine D1 receptor actions in layers V-VI rat prefrontal cortex neurons *in vitro*: modulation of dendritic-somatic signal integration. *J. Neurosci.* 16, 1922–1935 (1996).
59. Zheng, P., Zhang, X. X., Bunney, B. S. & Shi, W. X. Opposite modulation of cortical N-methyl-D-aspartate receptor-mediated responses by low and high concentrations of dopamine. *Neuroscience* 91, 527–535 (1999).
60. Rao, S. G., Williams, G. V. & Goldman-Rakic, P. S. Destruction and creation of spatial tuning by disinhibition: GABA(A) blockade of prefrontal cortical neurons engaged by working memory. *J. Neurosci.* 20, 485–494 (2000).
61. Gulledge, A. T. & Jaffe, D. B. Dopamine decreases the excitability of layer V pyramidal cells in the rat prefrontal cortex. *J. Neurosci.* 18, 9139–9151 (1998).

62. Diesmann, M., Gewaltig, M. O. & Aertsen, A. Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402, 529–533 (1999).
63. Riehle, A., Grün, S., Diesmann, M. & Aertsen, A. Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* 278, 1950–1953 (1997).
64. Domjan, M. & Burkhard, B. *The Principles of Learning and Behavior* 3rd ed. (Brooks/Cole, Pacific Grove, California, 1993).
65. Lisman, J. E. & Idiart, A. P. Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science* 267, 1512–1515 (1995).
66. O'Reilly R. C., Braver T. S. & Cohen J. D. in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (eds. Miyake, A. & Shah, P.) 375–411 (Cambridge Univ. Press, Cambridge, 1999).
67. Haj-Dahmane, S. & Andrade, R. Ionic mechanism of the slow afterdepolarization induced by muscarinic receptor activation in rat prefrontal cortex. *J. Neurophysiol.* 80, 1197–1210 (1998).
68. Broersen, L. M. *et al.* Effects of local application of dopaminergic drugs into the dorsal part of the medial prefrontal cortex of rats in a delayed matching to position task: comparison with local cholinergic blockade. *Brain Res.* 645, 113–122 (1994).
69. Wilson, H. R. & Cowan, J. D. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* 13, 55–80 (1973).
70. Amari, S. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* 27, 77–87 (1977).
71. Goodridge, J. P., Dudchenko, P. A., Worboys, K. A., Golob, E. J. & Taube, J. S. Cue control and head direction cells. *Behav. Neurosci.* 112, 749–761 (1998).
72. Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* 16, 2112–2126 (1996).
73. Camperi, M. & Wang, X.-J. A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *J. Comput. Neurosci.* 5, 383–405 (1998).
74. Romo, R., Brody, C. D., Hernández, A. & Lemus, L. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399, 470–473 (1999).
75. Seung, H. S., Lee, D. D., Reis, B. Y. & Tank, D. W. Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron* 26, 259–271 (2000).
76. Milner, B. & Petrides, M. Behavioural effects of frontal-lobe lesions in man. *Trends Neurosci.* 7, 403–407 (1984).
77. Funahashi, S., & Kubota, K. Working memory and prefrontal cortex. *Neurosci Res.* 21,1–11 (1994).

Viewpoint • Facilitating the science in computational neuroscience

'Computational neuroscience' means different things to different people, but to me, a defining feature of the computational approach is that the two-way bridge between data and theory is emphasized from the beginning. All science, of course, depends on a symbiosis between observation and interpretation, but achieving the right balance has been particularly challenging for neuroscience. Here I discuss some of the difficulties facing the field, and suggest how they might be overcome.

The first problem is that quantitative experiments are generally difficult and time consuming, and it is simply not possible to do all the experiments that one might think of. Nor is it possible to publish all the data that any given experiment generates. Given that so much must be excluded, it is essential that the experiments should be guided by theory, if they are to yield more than an arbitrary collection of unfocused facts. Conversely, theory needs to be informed by experimental data: too many theoretical papers present hypotheses that are incompatible with known facts about biology, and this problem is exacerbated by the difficulty theorists face in keeping up with a large and ever-expanding experimental literature.

How might the situation be improved? One step would be to ensure that theoretical papers are reviewed by experimentalists. This would help theoreticians not only to keep current with the experimental literature, but also to develop a better appreciation of how data are presented. Theoreticians are often tempted, for example, to extract quantitative information from representative examples of 'raw' data, failing to realize that 'representative' usually means 'best typical', thus compromising any practical utility.

Theoreticians also need to improve the presentation of their own models. It is taken for granted that experimental papers should contain sufficient information for others to replicate the results, but unfortunately, much theoretical work neglects this basic principle. Attempts to reproduce published computer models often fail, and it is difficult to know whether such failures reflect something profound, or whether they arise simply because the documentation of models with many parameters is naturally prone to error.

Experimental neuroscientists are not likely to pay serious attention to theoretical models until this problem is resolved. One solution is to develop a standard format for expressing model structure and parameters, and indeed this goal is evident in various neuroscience database projects currently underway. Supplying model source code is usually not enough. The format should be efficient and concise, yet allow a level of generic expression readable by humans and readily translatable for different simulation and evaluation tools. These requirements suggest exploiting programming languages oriented toward symbolic as well as numeric relations. It will be encouraging if such a standard is adopted at the publication level, because this will facilitate a more thorough review process as well as provide an accessible database for the reader. Eventually, this approach can contribute to a seamless database covering the entire field of neuroscience.

Finally, it is vital for this young field that the scientific and funding environment allow many interdisciplinary flowers to bloom. Support is needed for the marriage of theory and experiment at *all* levels of neuroscience, ranging from the biophysical basis of neural computation, to the neural coding of the organism's external and internal worlds, all the way up to the mysterious but (we assume) concrete link between brain and mind. Progress at the first level in particular will be essential if any rational medical therapeutics are to emerge from all this work. Core neuroscience courses should include a theoretical component, demonstrating its fundamental relevance to experimental neuroscience. At the same time, an ongoing critical examination of this marriage is necessary for the evolution of computational neuroscience. Perhaps we could learn lessons from physics, in which there is a more mature liaison between theory and application. As neuroscientists we may not avoid the occasional wild goose chase, but we can at least hope that a theory or two may be falsified in the process, clearing the path a bit for the next go-around and making it all worthwhile.

LYLE BORG-GRAHAM
Unité de Neurosciences Intégratives et Computationnelles,
Institut Federatif de Neurobiologie Alfred Fessard, CNRS,
Avenue de la Terrasse, 91198 Gif-sur-Yvette, France
e-mail: lyle@cogni.iaf.cnrs-gif.fr