European Journal of Neuroscience, Vol. 18, pp. 2011–2024, 2003

Retrospective and prospective persistent activity induced by Hebbian learning in a recurrent cortical network

Gianluigi Mongillo,[1] Daniel J. Amit[2,3] and Nicolas Brunel[4]
[1]Dipartimento di Fisiologia Umana, Università di Roma La Sapienza, Rome, Italy
[2]INFM, Dipartimento di Fisica, Università di Roma La Sapienza, Rome, Italy
[3]Racah Institute of Physics, Hebrew University, Jerusalem, Israel
[4]CNRS, Neurophysique et Physiologie du Système Moteur, Université René Descartes, Paris, France

Keywords: integrate-and-fire neuron, network model, perirhinal cortex, prefrontal cortex

Abstract

Recordings from cells in the associative cortex of monkeys performing visual working memory tasks link persistent neuronal activity, long-term memory and associative memory. In particular, delayed pair-associate tasks have revealed neuronal correlates of long-term memory of associations between stimuli. Here, a recurrent cortical network model with Hebbian plastic synapses is subjected to the pair-associate protocol. In a first stage, learning leads to the appearance of delay activity, representing individual images ('retrospective' activity). As learning proceeds, the same learning mechanism uses retrospective delay activity together with choice stimulus activity to potentiate synapses connecting neural populations representing associated images. As a result, the neural population corresponding to the pair-associate of the image presented is activated prior to its visual stimulation ('prospective' activity). The probability of appearance of prospective activity is governed by the strength of the inter-population connections, which in turn depends on the frequency of pairings during training. The time course of the transitions from retrospective to prospective activity during the delay period is found to depend on the fraction of slow, *N*-methyl-D-aspartate-like receptors at excitatory synapses. For fast recurrent excitation, transitions are abrupt; slow recurrent excitation renders transitions gradual. Both scenarios lead to a gradual rise of delay activity when averaged over many trials, because of the stochastic nature of the transitions. The model reproduces most of the neuro-physiological data obtained during such tasks, makes experimentally testable predictions and demonstrates how persistent activity (working memory) brings about the learning of long-term associations.

Introduction

Neurophysiological experiments have established persistent delay activity as the main candidate for a neuronal substrate of working memory (Fuster & Alexander, 1971; Funahashi *et al*., 1989; Fuster, 1995; Goldman-Rakic, 1995). Persistent delay activity was first discovered in prefrontal cortex (PFC), and later in inferotemporal (IT) cortex (Fuster & Jervey, 1981; Miyashita & Chang, 1988) and other areas of the temporal lobe (Nakamura & Kubota, 1995). Miyashita (1988) found links between persistent activity and long-term associative memory: if training in the delay-match-to-sample task is performed with a fixed sequence of sample images, single cells in the temporal lobe show elevated delay activity, following presentations of several images that are neighbours in the sequence. Thus, correlations between delay activity patterns reflect temporal associations between stimuli.

Sakai & Miyashita (1991) and Naya *et al*. (2001, 2003) used a pair-associate task to investigate further links between associative memory and persistent activity. Images shown to the monkey were divided into fixed pairs (Fig. 1A). A trial consisted of the presentation of one image of a pair (the cue or predictor), followed by a delay, and finally by a test (or choice) stimulus that includes the pair-associate of the cue together with a distractor. The monkey was rewarded for touching the 'pair-associate' of the cue. Following long training, some neurons, visually responsive for a particular picture, showed increasing activity in the delay period, preceding the presentation of that picture as a test stimulus, i.e. when the monkey expected that picture to be shown as test (prospective activity, see also Rainer *et al*., 1999; Fuster, 2001). These neurons have been termed 'pair-recall' neurons. For some neurons, visual responses to the pair-associates became highly correlated ('pair-coding' neurons).

More recently, Erickson & Desimone (1999) devised a task that allowed to record during the learning of new pairs. The task associates a fixed test stimulus to a go/no-go choice. In 85% of the trials, the test stimulus was preceded by its pair-associate ('predictor') stimulus (Fig. 1B). Such protocol reduces the learning phase (monitored by the monkey's performance level) to one or two sessions. It was found that the delay activity between predictor and choice presentations in perirhinal (PRh) cortex changed, during learning, from representing purely the predictor (retrospective activity) to representing both predictor and choice (prospective activity). With novel stimuli there was no similarity in visual responses of paired stimuli, and inter-stimulus delay activity was purely retrospective. With familiar stimuli, PRh neurons showed high correlation of visual responses to consistently paired stimuli, and the delay activity was correlated with both the predictor and the choice stimuli.

Possible mechanisms for persistent activity have been explored by theoretical modelling (Amit, 1995; Durstewitz *et al*., 2000; Wang, 2001). The main candidate is the reverberation mechanism through excitatory feedback (Hebb, 1949). The synaptic structure sustaining persistent activity can be a consequence of Hebbian plasticity induced

Correspondence: Dr N. Brunel, as above.
E-mail: brunel@biomedicale.univ-paris5.fr

Received 23 May 2003, revised 10 July 2003, accepted 14 July 2003

doi:10.1046/j.1460-9568.2003.02908.x

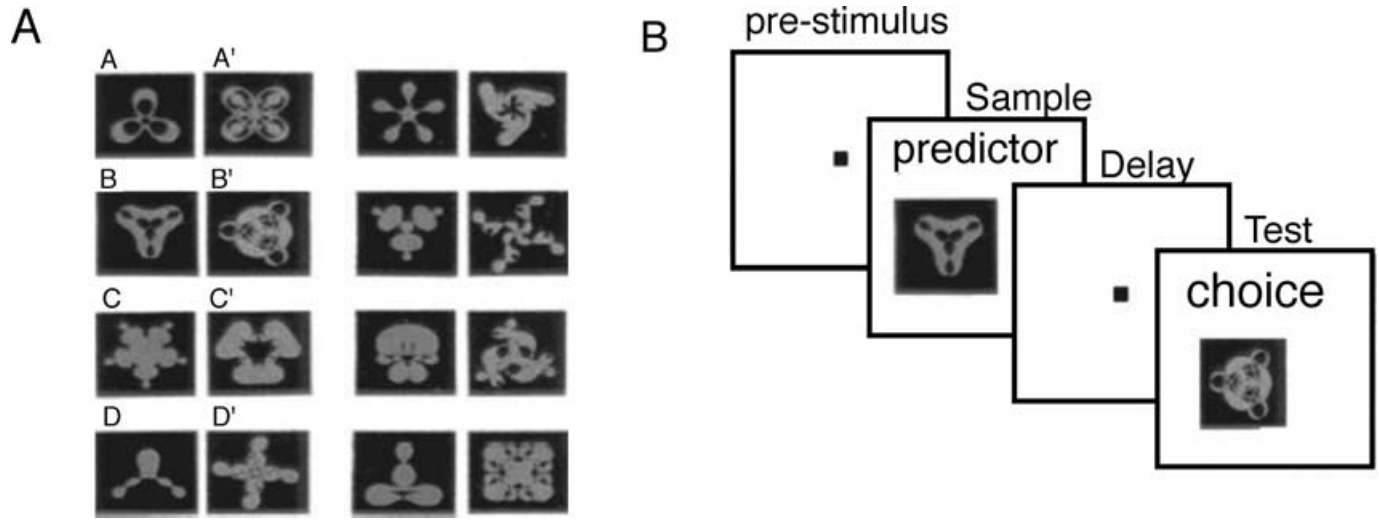

FIG. 1. Pair-associate task (adapted from Naya *et al*., 1996; Rainer *et al*., 1999). (A) A set of images is divided into fixed pairs (associates) (A, A′) (B, B′), etc. (adapted from Naya *et al*., 1996). (B) Protocol of the task: prestimulus interval, sample presentation, delay period and test (choice) presentation.

by stimuli (Amit, 1995; Amit & Brunel, 1997b). The link between persistent activity and stimulus–stimulus associations (Miyashita, 1988) has been explored in several studies (Griniasty *et al*., 1993; Amit *et al*., 1994; Brunel, 1996). However, these studies dealt only with stationary properties in the delay period, using mean-field approaches. Temporal dynamics during the delay period has not been explored by modelling studies.

The pair-associate paradigm provides a unique terrain for studying the inter-play between learning and persistent activity. We use it to investigate the evolution of persistent activity during learning in a pair-associate task in a model cortical network with plastic synapses. We find that learning naturally leads first to the appearance of 'retrospective' persistent activity, and later to the appearance of 'prospective' activity.

Materials and methods

The model network

We model a 'cortical module' of an area of the temporal lobe where selective persistent activity related to objects is observed. The model is composed of N_E pyramidal cells and N_I inhibitory inter-neurons. Each neuron receives, on average, C_E synaptic contacts from excitatory neurons and C_I from inhibitory neurons inside the network (selected at random), and C_{ext} excitatory synaptic contacts representing external afferents (Amit & Brunel, 1997b). The external afferents are activated independently by a Poissonian process, with rate ν_{ext}. The current resulting from the activation represents both noise from the rest of the cortex as well as selective afferents due to the presentation of stimuli. Excitatory neurons in the network are assumed to be selective to a discrete set of p external stimuli (representing the images or objects shown in the experiments). To the p stimuli correspond p subpopulations, each consisting of fN_E excitatory neurons, where f ($f << 1$) is the 'coding level'. For the sake of simplicity, we assume subpopulations are non-overlapping, i.e. all neurons in a given population respond to a single stimulus. Stimuli are organized in $p/2$ associated pairs: Stimulus (A, A′) (B, B′), ... In our case, $p = 16$ stimuli are divided in eight pairs. The presentation of a stimulus is simulated by selectively increasing the external rates afferent to the corresponding population, $\nu_{ext} \rightarrow (1 + \lambda)\nu_{ext}$, where λ is the 'contrast' of external stimuli. The architecture of the model is shown in Fig. 2.

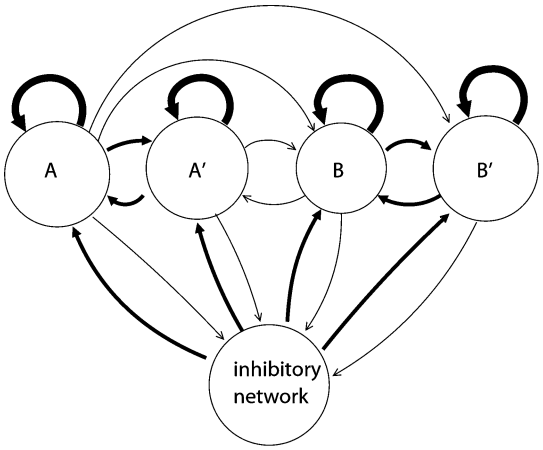

FIG. 2. Architecture of the model network. The network is composed of a large number of excitatory neurons and inhibitory neurons. Circles denote functional populations, labelled by the objects they encode. Arrows connecting populations are directional synaptic connections, whose thickness indicates their relative strength. Both excitatory and inhibitory neurons receive connections from 20% of excitatory neurons and 20% of inhibitory neurons (inhibitory to inhibitory connections not shown), as well as connections from outside the network (not shown). Inhibitory connections are stronger (on average) than excitatory connections, in order to render spontaneous activity stable (see, e.g. Amit & Brunel, 1997b). Disjoint populations of excitatory neurons A, B, ... represent the 'predictor' images and A′, B′, ... are their corresponding pair-associates. Following learning, connections within subpopulations are much stronger than average, while connections between pair-associate populations (e.g. A → A′) are only slightly stronger than average. The figure represents a network with a symmetric synaptic matrix. In the asymmetric scenario, connections from A to A′ are stronger than connections from A′ to A (see text).

The neurons of the network are leaky integrate-and-fire (IF) neurons. The state of a neuron is described by its depolarization $V(t)$, obeying the equation:

$$\tau_m dV(t)/dt = -V(t) + I(t) \quad (1)$$

where $I(t)$ is the total afferent current (in units of V) due to spikes arriving from presynaptic neurons; τ_m is the membrane time constant. When $V(t)$ reaches a threshold θ, the neuron emits a spike and V is reset to V_r, following a refractory period τ_{arp}. The synaptic current $I(t)$ is the

sum of individual post-synaptic currents induced by the C_E excitatory synapses and the C_I inhibitory synapses. Individual post-synaptic currents obey the equation,

$$\tau_s dI_s(t)/dt = -I_s(t) + \tau_m J \sum_k \delta(t - t_k - \delta_s) \quad (2)$$

where τ_S is the decay time constant of the synaptic current; J is proportional to the total charge transmitted by a single spike across the synapse (its efficacy, in mV units) and δ_S is the associated latency; t_k is the time of the synaptic activation, due to the k-th presynaptic spike, and spikes are described by instantaneous current injections. Eqn 2 implies that, upon emission of a presynaptic spike, the post-synaptic current has, following a delay δ_S, an instantaneous jump proportional to the efficacy, followed by an exponential decay with a time constant τ_S. Dependence on the neurotransmitter involved is taken into account by using different τ_S. The inhibitory synapses produce a fast inhibitory current mimicking the γ-aminobutyric acid (GABA) current ($\tau_S = 5$ ms). The recurrent excitatory synapses have both a fast ($\tau_S = 2$ ms) and a slow ($\tau_S = 100$ ms) component, corresponding, respectively, to AMPA and *N*-methyl-D-aspartate (NMDA) currents. A fraction x of the total charge is assumed to be transmitted by the slow component, and the remaining fraction $(1 - x)$ by the fast component. External excitatory synapses have only a fast component, with efficacy J_{Eext}, J_{Iext}, to excitatory/inhibitory neurons, respectively. The total current afferent on a neuron, $I(t)$ in Eqn 1, is the sum of the different components, each evolving with its own time constant. The voltage-dependence of NMDA is not modelled.

We have studied the behaviour of the model in two successive stages. First, the neural dynamics is studied during single trials, with a fixed, prestructured synaptic matrix. Second, once neuronal dynamics is well understood, the full learning scenario is implemented, to investigate the inter-play between neuronal and synaptic dynamics.

The protocols

We have simulated the pair-associate protocols with ordered pairs (Erickson & Desimone, 1999), i.e. the first member of a pair (A) is used only as a cue (predictor), while the second member appears only as a test (choice). The simulations reproduce, in the spirit of Amit (1998), 2 days of the experiment of Erickson & Desimone (1999), consisting of a series of 1000 trials (2 days of 500 trials each). Each trial consists of four intervals.

1. Prestimulus (1000 ms): no selective external inputs.
2. Cue (predictor) presentation (500 ms): a randomly chosen stimulus from the set of predictor cues (e.g. A) is shown to the network. The activation rate of the afferents to the neurons of the corresponding population is increased to $(1 + \lambda)\nu_{ext}$.
3. Delay interval (1000 ms): no selective external inputs.
4. Test (choice) presentation (500 ms): in 85% of trials, the pair-associate (A′) of the cue (A) is shown to the network ('valid' trials). In 15% of trials, another randomly chosen stimulus from the set of choice stimuli (e.g. C′) is presented ('invalid' trials). In the simulations, the fraction of 'valid' trials was either 100% or 85%.

Other protocols have used unordered pairs (Sakai & Miyashita, 1991; Naya *et al*., 1996). In these protocols, any of the stimuli of a pair (A or A′) can appear as a cue. The test stimulus is composed of the pair-associate of the cue image, together with a distractor image.

Network with prestructured synaptic matrix

The synaptic matrix is constructed at the beginning of the simulation and stays fixed thereafter. The process of building the synaptic matrix is done in two steps. First, for each excitatory neuron, we select the set of C_E excitatory presynaptic neighbours of that neuron, randomly and independently from neuron to neuron. This defines the set of functional synapses of the network, at which plasticity can take place. A similar procedure is done for other populations of synapses (inhibitory synapses on excitatory neurons, excitatory and inhibitory synapses on inhibitory neurons), but these synapses all have a fixed and equal efficacy J_{IE}, J_{EI} and J_{II}, respectively. Next, each existing excitatory synapse on an excitatory neuron J_{ij} (where j denotes the presynaptic neuron and i the post-synaptic neuron) is assigned one of two possible states, a potentiated (up) state with efficacy $J_{ij} = J_1$ and a depressed (down) state with efficacy $J_{ij} = J_0$. Structuring due to learning is expressed in shifting the proportion of synapses in the up and down states. In the final outcome of the training stage, the probability for a synapse to be in the up state depends on whether the protocol uses ordered or unordered pairs (see below).

In the general case, the structure in the resulting synaptic matrix is potentially asymmetric, with

$$\mathrm{Prob}(J_{ij=J_1}) = \begin{cases} 1 & \text{if } i,j \text{ in the same population} \\ a & \text{if } j\text{(postsynaptic) in } predictor \\ & \text{population (e.g. A)} \\ & \text{and } i \text{ (presynaptic) in } choice \\ & \text{population (e.g. A}') \\ a' & \text{if j in choice population (e.g. A}') \\ & \text{and i in predictor population (e.g. A)} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where a is a forward pair-learning parameter (strength of synapses whose presynaptic neuron is selective for a 'predictor', e.g. A, while the post-synaptic neuron is selective for a 'choice', e.g. A′), and a' is a backward pair-learning parameter (vice versa). If in training the pairs are of fixed order (first A, then A′), the resulting synaptic matrix may have $a \neq a'$ – asymmetric structuring. If the images within the pairs are presented at random, the resulting synaptic matrix will have $a = a'$ – symmetric structuring. The symmetry/asymmetry of the synaptic matrix depends not only on ordering of the pairs, but also on the symmetry/asymmetry of the learning dynamics (see below). Hence, in principle, even in the ordered pair case, the resulting inter-population synaptic structure may end up symmetric.

In the following, for the prestructured case, we will consider the two extremes: $a = a'$ (symmetric) and $a' = 0$ (fully asymmetric).

Network with learning dynamics

Plasticity is restricted to excitatory-to-excitatory synapses. The synaptic matrix is initialized by assigning to each existing excitatory-to-excitatory synapse in the connectivity scheme described above, the efficacy J_1 with probability 0.05, and J_0 otherwise, irrespective of the identity of pre- and post-synaptic neurons (tabula rasa). The learning process is implemented in a Hebbian, rate-dependent way between excitatory neurons only. Plasticity takes place only in existing synapses of the random connectivity arrangement. The average spike rate of every excitatory neuron is estimated as the ratio of the number of spikes emitted into a time window T divided by T. The time window slides by ½T increments, so that each trial is divided into overlapping bins of T ms. If in a window T both cells emit at a rate above a high threshold T_+, chosen to be lower than the rate of visual response, but higher than the rate in delay activity, and the synapse has efficacy J_0, its efficacy is potentiated to J_1, with probability p_+ [strong long-term potentiation (LTP) condition]; if the presynaptic cell emits at a rate below T_+ but above a low threshold T_a (lower than delay activity rate but higher than the rate in spontaneous activity), while the post-synaptic cell emits at a rate above T_+, the efficacy $J_0 \rightarrow J_1$ with probability p_w (weak LTP condition 1); in the opposite case, if the presynaptic cell emits at a rate above T_+, while the post-synaptic rate is

below T_+ but above T_a, the efficacy $J_0 \rightarrow J_1$, with probability $p_{w'}$ (weak LTP condition 2); if the rate of one of the two cells is above T_+ and that of the other cell is below T_a, $J_1 \rightarrow J_0$, with probability p_- (LTD condition); in the remaining cases, when none of the two cells emits at high rate, no change occurs. This plasticity dynamics is motivated in the Discussion. To summarize,

$$\text{Prob}(J_0 \rightarrow J_1) = \begin{cases} p_+ & \text{strong LTP condition} \\ p_w & \text{week LTP condition 1} \\ p_{w'} & \text{week LTP condition 2} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Prob}(J_0 \rightarrow J_1) = \begin{cases} p_- & \text{LTD condition} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The regions where plasticity takes place, in the plane of pre- and post-synaptic rates, are shown in Fig. 3.

If $p_{w'} \neq p_w$ and the ordering of the images within the pairs is fixed, this learning dynamics leads asymptotically to an asymmetric structure, Eqn 3 with $a \neq a'$. However, if during training the pairs are not ordered, i.e. each element of the pair is as often presented as predictor or as choice, the structuring will end up symmetric (Eqn 3, with $a = a'$). The simulation of the full learning dynamics was carried out in the fully asymmetric case ($p_{w'} = 0$), as in Erickson & Desimone (1999). The main goal has been to check that the dynamics indeed converges to the expected synaptic matrix.

Parameters of the network of IF neurons

The simulated network of IF neurons had the parameters given in Table 1.

The parameters related to the network architecture were chosen to be compatible with realistic cortical anatomy. Individual neuronal parameters and synaptic temporal parameters were chosen in accordance with known physiological data (McCormick *et al*., 1985; Mason *et al*., 1991; Markram *et al*., 1997). The synaptic latency was drawn randomly and independently from a uniform distribution, in an interval (δ) given in Table 1. The amplitudes of synaptic efficacies and external rates were chosen to obtain background 'spontaneous' activity at about 5 Hz for pyramidal cells and 10 Hz for inter-neurons. J_1 was chosen to ensure stable persistent activity (for fixed network structure) or to lead to such stable activity in the learning process (for evolving network structure). The relationship between J_0 and J_1 is chosen so that spontaneous activity is unchanged as J_1 is varied (Amit & Brunel, 1997b). The contrast of external stimuli, λ, was chosen to render the visual response, at the beginning of the learning process, higher than the high learning threshold T_+. In simulations with prestructured synaptic matrix, λ was chosen to produce a visual response of about 80 Hz. The value of T_+ ensures that synaptic modifications occur only during visual presentations, and not during delay activity or spontaneous activity, for structural stability reasons (Amit & Mongillo, 2003). The low learning threshold, T_a, was chosen to be higher than spontaneous activity, but otherwise as low as possible to allow synaptic modifications in the initial part of the interval of the choice stimulus presentation. Finally, the synaptic transition probabilities, p_+, p_w (recall that $p_{w'} = 0$), were chosen to be low, so that learning occurs gradually over the course of many presentations; the weak LTP probability was chosen to be lower than the high LTP probability again for stability reasons. Too high learning probabilities (in particular

TABLE 1. Parameters of the network of IF neurons

Number of excitatory neurons	N_E	8000
Number of inhibitory neurons	N_I	2000
Number of recurrent excitatory connections per neuron	C_E	1600
Number of external excitatory connections per neuron	C_{ext}	1600
Number of inhibitory connections per neuron	C_I	400
Coding level	f	0.05
Number of stimuli	p	16
Membrane time constant, pyramidal cells	τ_{mE}	20 ms
Membrane time constant, inter-neurons	τ_{mI}	10 ms
Firing threshold (both types)	θ	20 mV
Reset membrane potential, pyramidal cells	V_{rE}	10 mV
Reset membrane potential, inter-neurons	V_{rI}	15 mV
Refractory period (both types)	τ_{ARP}	2 ms
Average E → E efficacy	J_{EE}	0.05 mV
E → I efficacy	J_{IE}	0.11 mV
I → E efficacy	J_{EI}	0.15 mV
I → I efficacy	J_{II}	0.26 mV
External E → E efficacy	J_{Eext}	0.055 mV
External E → I efficacy	J_{Iext}	0.1 mV
Potentiated E → E efficacy	J_1	3.2 J_{EE}
Depressed E → E efficacy	J_0	$(J_{EE} - fJ_1)/(1-f)$
Synaptic decay type: fast excitation (AMPA-like)	τ_{AMPA}	2 ms
Synaptic decay type: slow excitation (NMDA-like)	τ_{NMDA}	100 ms
Synaptic decay type: fast inhibition (GABA-like)	τ_{GABA}	5 ms
Fraction of slow excitatory current	x	0.05–0.30
Latency (transmission delay)	δ	0.5–3.5 ms
Background external rates	ν_{ext}	15 Hz
Contrast of external stimulus	λ	0.7
High learning threshold	T_+	65 Hz
Low learning threshold	T_a	15 Hz
LTP probability	P_+	0.007
Weak LTP probability	p_w	0.0035
LTD probability	p_-	0.007
Learning bin	T	100 ms
Pair-learning parameter (for fixed synaptic matrix)	A	0–0.04

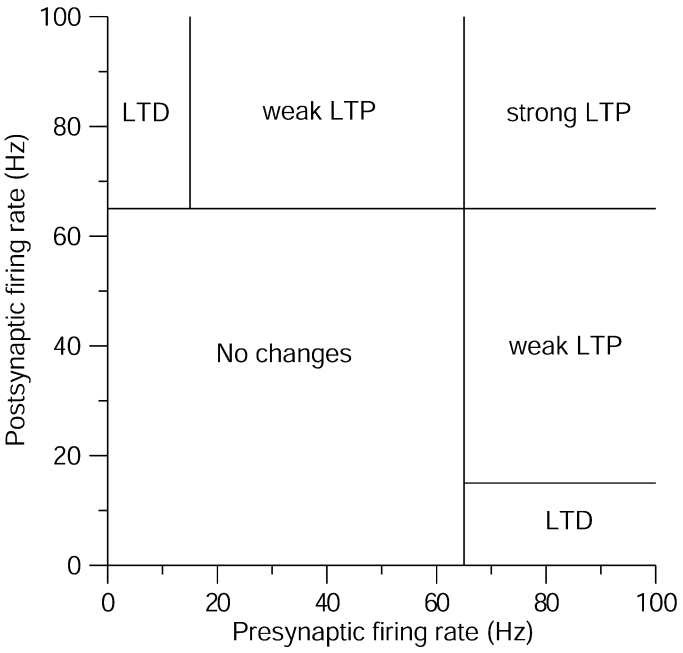

FIG. 3. Regions of synaptic transitions in the space of pre- and post-synaptic rates. LTP occurs when both pre- and post-synaptic rates are high, above a high threshold $T_+ = 65$ Hz; LTD occurs when one rate is high (above T_+) and one below a low threshold, $T_a = 15$ Hz; weak LTP occurs when one rate is high (above T_+) and one intermediate (between T_a and T_+); otherwise no change occurs.

too high weak LTP probability) lead to an epileptic state, when too many inter-population synapses become potentiated to J_1.

Simplified (population-rate) model

We also studied the network dynamics by using a simplified model. The full simulation of the network of spiking neurons, subject to the entire experimental protocol including the synaptic formation, is quite time consuming. For example, to run the full set of 1000 trials in the pair-associate paradigm takes several days on a fast workstation. By contrast, a 'mean-field' approach provides an explicit, complete and rapid picture of the attractor landscape of the network, hence the available stationary states of its dynamics. The large-scale simulation of the network of spiking neurons is used to confirm the results of the mean-field model and to explore the role of transitions between attractors due to the intrinsic fluctuations related to the finite size of the system, absent in the mean-field approach.

Note that the simplified model is not meant as an approximation to the full simulation (as in, e.g. Amit & Brunel, 1997a), but rather as a simple tool for a qualitative study of possible stationary network states as the synaptic matrix is varied.

Excitatory neurons in the simplified network are chosen to have an *f–I* curve of the form:

$$\phi(I) = \begin{cases} 0 & I < 0 \\ V_c(I/I_c)^2 & 0 < I < I_c \\ 2V_c(I/I_c - 3/4)^{1/2} & I > I_c \end{cases} \quad (5)$$

giving the firing rate ν vs. the mean input current I, shown in Fig. 4A. I_c can be thought of as a threshold current (again in mV), while ν_c is the typical firing rate of cells at this threshold current in presence of realistic noise. We use for the sake of illustration $I_c = 20$ mV and $\nu_c = 10$ Hz. The *f–I* curve of Eqn 5 is chosen for its simplicity, and because it reproduces the qualitative features of spiking neurons in the presence of noise. It has a convex subthreshold region (for $I < I_c$), mimicking the noise-driven region in spiking neurons (note that the power law behaviour is a good approximation of the *f–I* curve in a wide parameter range for many neuronal models (Hansel & van Vreeswijk, 2002). It has a suprathreshold region (for $I > I_c$), with a square root dependence on the input current, as expected for type I neurons (Ermentrout, 1996).

A state of the network is described by the mean rate of the p non-overlapping excitatory subpopulations, each selective for a particular stimulus, and the mean rate of the non-selective inhibitory population. For simplicity, the activity of inhibitory population is assumed to depend linearly on the average activity of the excitatory populations. The fraction of excitatory neurons selective to a given image, the 'coding level', is chosen $f = 1/p$. In other words, there are fN neurons coding for each stimulus, and every excitatory neuron finds itself in one of the p, non-overlapping subpopulations. The input current to a neuron in population α, i.e. selective to stimulus number α, is denoted by I_α, and the mean spike rate in this population is ν_α.

The total synaptic strength from all neurons in population β to a single neuron in population α is $J_{\alpha\beta}$ (in mV·s units). The input current to population α is

$$I_\alpha = I_{\text{ext}} + I_{\text{st}} + \sum_\beta J_{\alpha\beta}\nu_\beta - J_1 \sum_\beta \nu_\beta \quad (6)$$

where the first term on the right-hand side corresponds to the background (non-selective) external afferent current, the second term is the selective input due to presentation of a stimulus, the third is the excitatory recurrent feedback, and the last term represents the inhibitory feedback, which we assume to be linearly proportional to the average activity in the excitatory network.

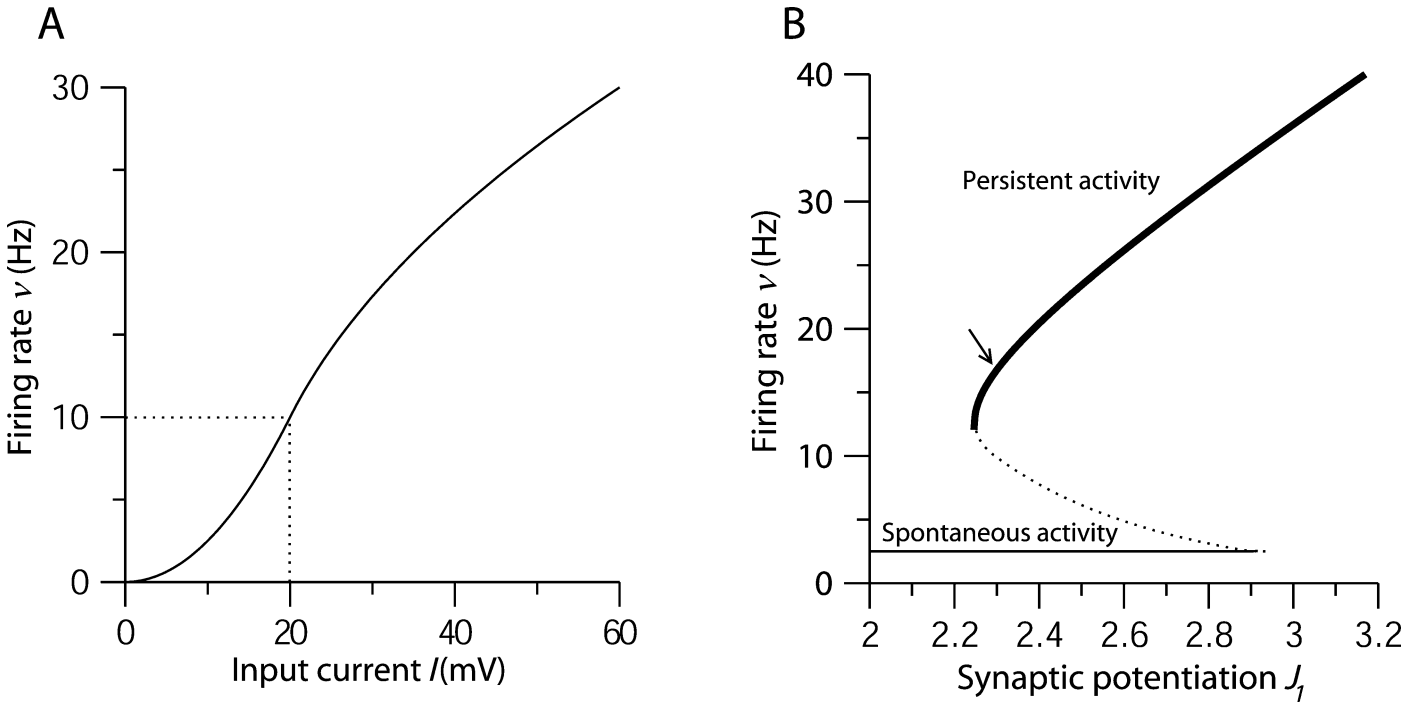

FIG. 4. The simplified model: (A) *f–I* curve, $\phi(I)$ of an excitatory neuron, Eqn 5, with $I_c = 20$ mV, $\nu_c = 10$ Hz. (B) Bifurcation diagram for the average spike rates in spontaneous and persistent delay activity states, as a function of the strength of potentiated synapses, J_1, in the absence of pair-associate learning: $a = 0$. Dotted line, boundary between the basins of the two stable states, when they coexist. For $J_1 < 2.24$ mV·s only spontaneous activity is stable; above 2.24 mV·s, both spontaneous activity and selective delay activity of each of the 20 subpopulations coexist. The arrow indicates the value of J_1 (= 2.3 mV·s) used in Figs 5 and 6.

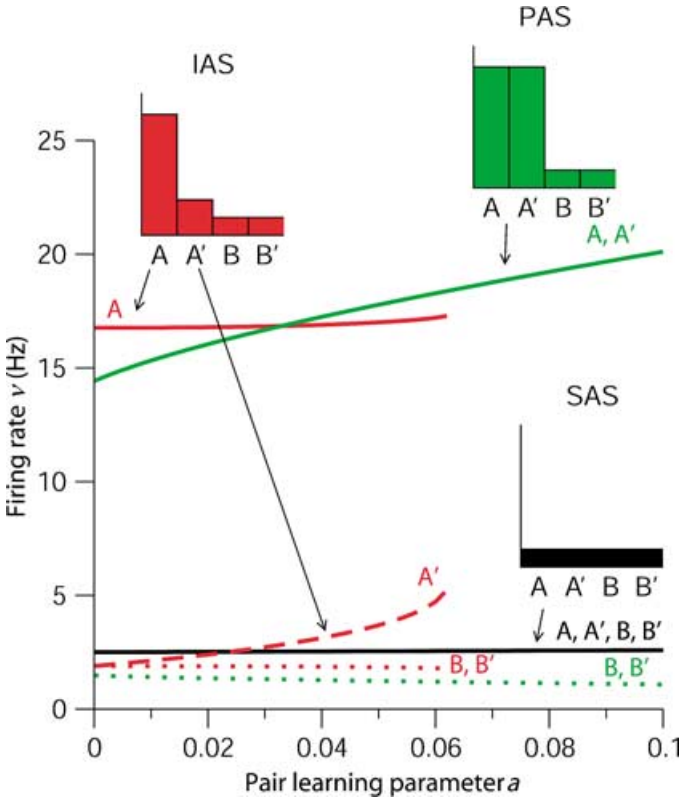

FIG. 5. Attractors in the simplified model with symmetric learning. Spike rates (arbitrary units) in stable network states vs. the pair-learning parameter, a. Three types of states are shown. (1) Spontaneous activity state (SAS, black line), in which all populations are in spontaneous activity. This state is identical to the unstructured spontaneous activity of Fig. 4B and exists in the entire range of $0 < a < 0.1$. (2) Individual attractor state (IAS, red lines), in which one of the populations (here A) emits at elevated rates (full red line), while the pair-associate (A′) emits at a rate which is slightly higher than spontaneous activity (dashed red line), due to increased connections between pair-associate populations. Other populations (dotted red line, B, B′, ...) emit at slightly lower rates than spontaneous activity, due to higher inhibitory activity caused by the delay activity. The IAS state is the analogue, in the pair-associate protocol, of the usual persistent activity state shown in Fig. 4B. It exists only at small values of a ($a < 0.06$). (3) Pair-attractor state (PAS, green lines), both populations of a pair (here A, A′) emit at elevated and equal delay rates (full green line). Other populations (B, B′, ...) emit at low rates (dotted green line). This state exists in the whole range of $0 < a < 0.1$. For $a < 0.06$, the three types of states coexist; for $a > 0.06$, only the spontaneous and pair-attractor states are stable. Insets show a schematic histogram of the rates in different populations in the network for the three types of states (SAS, IAS, PAS).

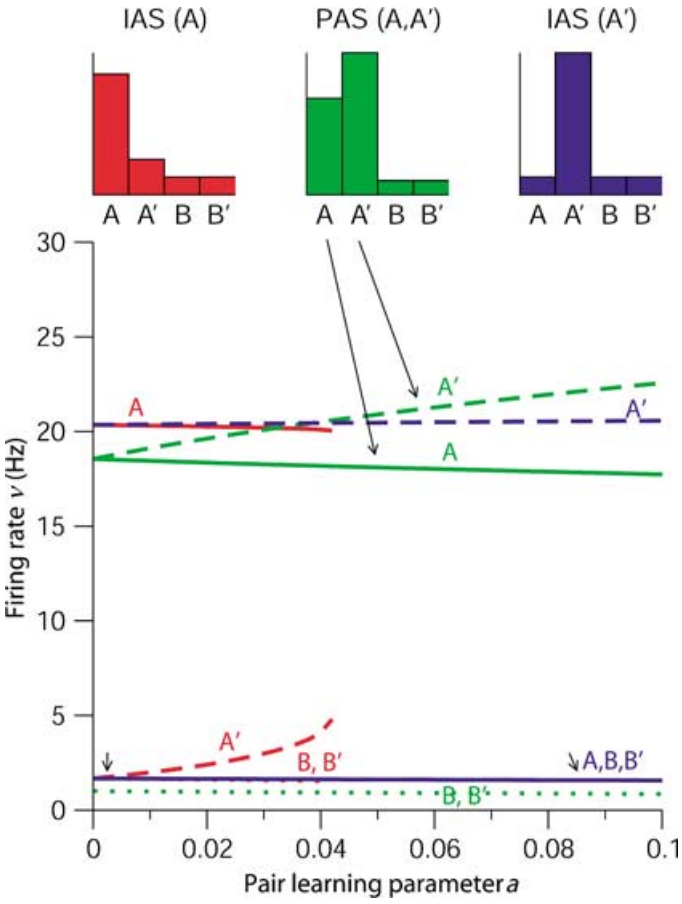

FIG. 6. Attractors in the simplified model with asymmetric learning of temporally ordered pairs. Rates in different network states vs. the pair-learning parameter a. Three types of selective states are shown [indicated schematically in the panels marked individual attractor states (IAS) (A), pair attractor states (PAS), IAS (A′)]: an IAS in which the first item of the pair (A) is active at an elevated rate, and the second item (A′) is weakly above spontaneous activity level. It exists only in the range $a < 0.05$. The IAS, in which the second item of the pair (A′) is active at elevated rates is stable in the range $0 \leq a \leq 1$. This is because elevated activity in A′ does not lead to increased activity in its pair-associate, which acts to destabilize the IAS in the symmetric scenario and the IAS (A) in the asymmetric scenario. Finally, the PAS exists in the range $a < 0.55$. The value of a where the PAS state becomes unstable strongly depends on the strength of inhibition. This PAS differs from the PAS of the symmetric learning case as the activity of A′ is stronger than the activity of A. When a is sufficiently strong, upon presentation of stimulus A, the network will make a transition either to the asymmetric PAS (both retrospective and prospective delay activity), or to the IAS of the second item (purely prospective delay activity).

The stationary average rates in the network are given by $\nu_\alpha = \phi(I_\alpha)$, Eqns 5 and 6. A simplified spike rate dynamics is used (Wilson & Cowan, 1972; Ermentrout, 1998)

$$\tau \mathrm{d}\nu_\alpha/\mathrm{d}t = -\nu_\alpha + \phi(I_\alpha) \tag{7}$$

where τ is a time constant associated with the rate dynamics.

The synaptic matrix incorporating the effect of pair-learning is expressed as:

$$J_{\alpha\beta} = \begin{cases} J_1 & \text{if } \alpha = \beta, \text{intra-population} \\ J_0 & \text{if } \alpha \neq \beta, \text{not pair-associates} \\ J_{0+a(J_1-J_0)} & \text{if } \alpha \text{ choice and } \beta \text{ predictor} \\ J_{0+d(J_1-J_0)} & \text{if } \alpha \text{ predictor and } \beta \text{ choice} \end{cases} \tag{8}$$

a, a' are the pair-learning parameters, introduced in Eqn 3. This synaptic matrix corresponds to an average over individual synapses whose weight is given by J_1 or J_0 according to Eqn 3.

The stationary states of the network can be represented as a 'bifurcation diagram' (Fig. 4B), where the stable rates are shown as a function of the potentiation strength J_1, for $a = a' = 0$ and $J_0 = 1 - f(J_1 - 1)/(1 - f)$ (Amit & Brunel, 1997b; Brunel, 2000b). Two types of attractor states are shown in Fig. 4B: (i) unstructured spontaneous activity, for which all populations have spontaneous activity levels (thin horizontal line); (ii) selective delay activity states, in which one population (the one that last received selective external inputs) has elevated activity (thick horizontal line), while all other populations remain close to spontaneous activity levels. Other types of states exist, as we will see later.

The spontaneous activity branch corresponds to the horizontal line in Fig. 4B. The selective activity branch (solid bold curve) starts at $J_1 = 2.24$ mV·s. Qualitative results, however, do not depend on the precise value of J_1, provided it is in the bi-stable range shown in Fig. 4B ($2.24 < J_S < 2.92$).

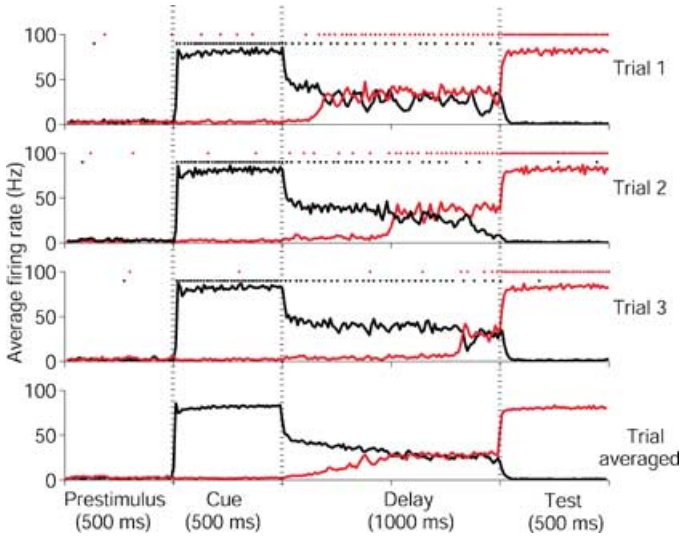

FIG. 7. Random transitions to prospective activity in the delay interval. Time course of population-averaged activity, with prestructured synaptic matrix, in the fully asymmetric condition with $a = 0.02$ and $x = 0.05$. The remaining parameters are reported in Table 1. The epochs of the trial are indicated in the bottom panel (prestimulus: 0–500 ms; cue presentation: 500–1000 ms; delay period: 1000–2000 ms; test presentation: 2000–2500 ms). Black curve: average rate in the predictor population. Red curve: average rate in the choice population. Rates are sampled in bins of 10 ms. Top three panels: single-trial examples of transition at different times during the delay period. At the top of each panel, we present a spike raster of one representative cell belonging to the predictor population (black) and one to the choice population (red). Note that retrospective activity can either persist (Trials 1 and 3) or die out (Trial 2). Bottom: predictor and choice population activity averaged over 100 trials (PSTH). The average delay activity in the pair-associate population shows a continuously increasing activity during the delay period.

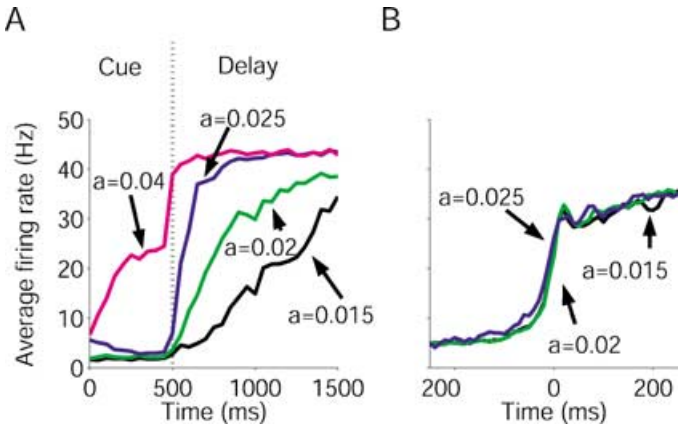

FIG. 8. Effect of pair-learning parameter a on prospective activity. Network with fully asymmetric prestructured synaptic matrix. (A) Temporal evolution of averaged prospective activity. Curves are trial-averaged activity of choice population during the cue presentation and the delay interval. The increase of the emission rate is due to the fact that the probability of having made a transition to a prospective state increases with the passage of time. It becomes steeper as a increases due to the decrease in the average lifetime of the IAS. For $a = 0.04$, transition occurs during cue presentation. (B) Temporal evolution of 'prospective' activity, synchronized on transition time ($t = 0$), defined by the first time at which population activity, averaged over 10-ms bin, exceeds 50% of its full increase and remains higher until the end of the delay period. The time course is unaffected by a. $x = 0.05$, other parameters are indicated in Materials and methods.

From a computational point of view, the attractor dynamics gives to the network properties similar to a winner-take-all network (e.g. Ermentrout, 1992), when attractors represent single images. This computational property is here a by-product of the dynamics of a recurrent network of excitatory and inhibitory neurons with Hebbian learning of discrete stimuli.

The parameters of the simplified network are: number of learned stimuli: $p = 20$; inhibitory efficacy: $J_I = 1.2$ mV·s; potentiated excitatory efficacy: $J_S = 2.3$ mV·s; background external inputs: $I_{ext} = 10$ mV. The stationary states of the simplified network were investigated by integrating numerically Eqns 5–8 using the Euler method with an integration time step $dt = 0.01\tau$, with various types of initial conditions (0, 1 or 2 subpopulations active at 20 Hz, the others at 2.5 Hz), until a stable fixed point was reached. The pair-learning indices a and a' were varied systematically. Note that the parameters of the simplified mean-field model are rather different than those of the spiking network model. This reflects the fact that the simplified model is not intended to describe quantitatively the dynamics of the network model, as mentioned above, but rather to explore qualitatively, in the simplest possible setting, the stationary states of the network as a function of possible synaptic structuring. Notice that in Eqn 6, J is in mV·s and not in mV as in the full network of IF neurons, as τ_m has been absorbed in the synaptic strengths.

Results

Simplified (population-rate) model

We start by analysing the stationary states of the simplified model. The synaptic matrix of the network, Eqn 8, is characterized by two pair-learning parameters (a and a'), which represent the strength of the connection ($A \rightarrow A'$ and $A' \rightarrow A$) between populations corresponding to pair-associate stimuli. Figures 5 and 6 show how the stable stationary states of the network (attractors) change as the pair-learning parameters vary. Figure 5 describes the case with symmetric synaptic matrix ($a = a'$), which would be obtained, for example, if training takes place with no particular order within the pairs (Sakai & Miyashita, 1991). Figure 6 describes the case with matrix, which would correspond to training with pairs of fixed order (Erickson & Desimone, 1999), and an asymmetric learning dynamics (see Materials and methods). For illustration we show in Fig. 6 the extreme case $a' = 0$.

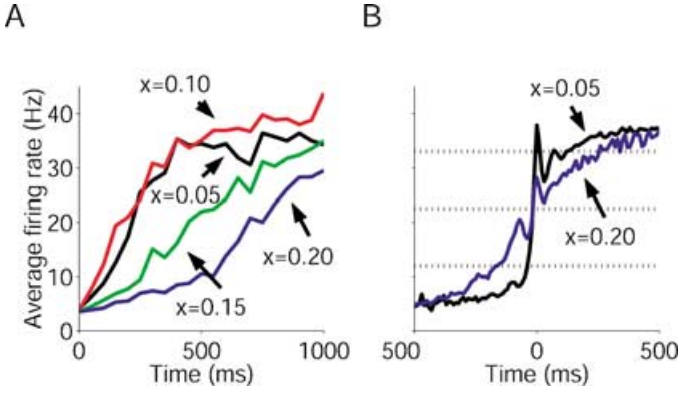

FIG. 9. Effect of synaptic kinetics on prospective activity. Network with fully asymmetric prestructured synaptic matrix for different values of x. (A) Temporal evolution of trial-averaged prospective activity for four values of the fraction of the slow currents x. As x increases, the increase in delay rate becomes slower. (B) Dynamics of the transition as a function of the fraction of the slow currents. Curves are trial-averaged activity of the choice population, synchronized at the transition time ($t = 0$). Horizontal dotted lines: 20%, 50% and 80% of the full rate increase. Larger x leads to slower transitions. For $x = 0.05$, the transition takes place in about 100 ms; for $x = 0.20$ it takes place in about 500 ms. Transition time is defined as in Fig. 8. $a = 0.025$, other parameters as in Fig. 8.

In both cases, the analysis reveals the existence of three types of attractor states.

1. Unstructured spontaneous activity state (SAS), in which all populations are active at low rates (black curve and black inset in Fig. 5). The spontaneous spike rates are unaffected by the values of a and a'.
2. Individual attractor states (IAS), one for each stimulus, in which a single population is active at elevated rate (red curves and red inset in Fig. 5). The delay period activity of the cue population (A) remains almost constant as a is varied. Because of the enhanced synapses connecting pair-associate populations, this persistent activity enhances the activity in the pair-associate population A′ (dashed red curve), above spontaneous rates. This rate increases with a and a'.
3. Pair attractor states (PAS), in which both populations of a given pair (A and A′) have elevated spike rates (green curves and green inset in Fig. 5). The pair state co-exists with the individual state at low values of a and a', down to $a = a' = 0$. At $a = 0$, all possible pair states [e.g. (A, A′) (A, B) (A, B′) (A′, B) (A′, B′) (B, B′)] are attractors and are equivalent. These states can be interpreted as simultaneous working memory of two items. As a increases, the basins of attraction of the learned pairs [e.g. (A, A′) and (B, B′)] expand, while the basins of attraction of other pairs shrink.

Symmetric case ($a = a'$)

Figure 5 shows that both types of states (individual and pair) co-exist until $a = 0.06$. For $a > 0.06$, the IAS are no longer stationary states of the dynamics and the PAS remain the only stable selective states of the system. This implies that for $a < 0.06$, if the network is stimulated by a single stimulus of a pair, it will end up in an IAS, while to reach a PAS it should be stimulated simultaneously by the two stimuli of the corresponding pair. Above $a = 0.06$, stimulating the network by any of the individual stimuli of a pair would lead to a pair stationary state, corresponding to the pair of which the stimulus is a member. For these values of a, the network is not able to maintain working memory of a single item of a pair.

Depending on the level of pair-learning, we can expect two types of 'prospective' delay activity in a given trial of a pair-associate task: if a is small, and A is shown as a cue, the network settles in an IAS, the individual predictor state, and the activity of neurons in population A′ will be only slightly enhanced compared with baseline – weak prospective delay activity. If a is large enough, and A is shown as a cue, the network settles in the PAS and the delay activity of neurons in population A′ will be elevated – strong prospective delay activity.

Asymmetric case ($a' = 0$)

From Fig. 6, one can read two qualitative differences in the structure of the state compared with the symmetric case. (i) The IAS corresponding to the predictor stimulus (A) becomes unstable at $a = 0.04$, while the one corresponding to A′ remains stable up to high values of a. As a consequence, when a becomes sufficiently strong ($0.05 < a < 0.55$), the network finds two selective states accessible: the IAS corresponding to A′ and the PAS. (ii) In the PAS, the rate of the choice stimulus is higher than that of the predictor. As in the symmetric case, until a critical value ($a = 0.05$), both IAS and PAS co-exist, in the sense explained above. Note that in the first state, retrospective activity is absent.

Network of spiking neurons: 'prospective' activity

Next, we turn to a microscopic simulation of a model of a 'cortical' network of IF neurons. This was done in two stages: (i) observing the neural dynamics in the network with a prelearned, fixed synaptic matrix incorporating pair-learning; (ii) observing the learning process in the microscopic simulation with coupled neural/synaptic dynamics.

Random transitions towards strong prospective activity occur in the delay period

As for the simplified network, the network of spiking neurons exhibits various steady states of selective delay (persistent) activity. In particular, a state in which an individual population has elevated delay rates while the remaining ones have much lower rates (IAS); and one in which two populations have elevated delay spike rates (PAS). The main difference from the simplified network is that the finite size of the microscopic network causes random fluctuations in the average activity of each population (Brunel & Hakim, 1999), and those provoke transitions between states. When such a transition occurs, the rate of neurons in the choice population rises – from weak to strong 'prospective' activity.

Depending on the degree of symmetry in the structuring and on the pair-learning parameter a, the retrospective activity can either persist all along the delay interval, or die out when the transition takes place. In the first case, the transition is between an IAS, corresponding to the predictor, and the corresponding PAS (i.e. the other active population is that of the paired choice). In the second case, the transition is between the IAS corresponding to the predictor and the IAS corresponding to the choice.

Examples of stochastic transitions can be viewed in Fig. 7, which shows the average activity of two populations, predictor and choice, during single trials, in a network with an asymmetric synaptic matrix. Neurons selective for the predictor stimulus (black curve) show high visual response during cue presentation and elevated delay activity when the stimulus is removed (retrospective activity). Neurons selective for the pair-associate stimulus (red curve) see their activity increase sharply at different instances during the delay period. These are spontaneous transitions induced by fluctuations and occur at random times during the delay period. The same neurons continue to be active as a visual response to the pair-associate (choice), when it is presented as test. As the transition takes place, the delay activity of the predictor population can either persist (PAS, Fig. 7, first and third panel) or die out (choice IAS, Fig. 7, second panel). If the transition does not occur, the retrospective delay activity persists all along the delay interval (i.e. until the presentation of the choice).

The situation is somewhat analogous to the escape rate of a random walker with a high threshold: the average time to escape is much longer than the time constants of the single neuron or of the synaptic dynamics, as the barrier is difficult to cross on these time scales. The distribution of escape times is close to exponential. The average delay activity of 'pair-associate' neurons becomes a slowly increasing function of time, with the slope at the origin inversely proportional to the average 'lifetime' of the individual attractor state. This slow rise of prospective activity is shown in the lower panel of Fig. 7.

Slope of rising prospective activity depends on pair-learning parameter

Figure 8A presents the time course of prospective activity during the delay period, averaged over 100 trials, for several values of a, in a network with asymmetric prestructuring. One observes a monotonic rise of the ′prospective′ activity, expressing the fact that the number of trials in which the transition has occurred increases with the passage of time. As a increases, the lifetime of the individual (A) attractor state decreases, and hence transitions to prospective activity occur earlier, leading to a higher slope of the trial-averaged activity. However, the dynamics of the transition itself, as revealed by synchronizing all rasters at the transition time, is quite sharp for any value of a (Fig. 8B).

The transition time is defined as the first time at which the activity of the subpopulation selective for A′ exceeds 50% of its full increase and remains higher than this level until the end of the delay period. The transition duration, as defined by the time it takes for population activity to rise from 20% to 80% of its full increase, is about 100 ms.

For small values of a, $a \leq 0.015$, the average transition time is much longer than the delay period. Hence, almost no transitions occur in the delay period, and the activity of the choice population remains approximately constant during the delay period. For a about 0.02, the average transition time becomes comparable to the delay period, hence the rise of prospective activity observed in Fig. 8. For larger values of a ($a \sim 0.04$), the transition to the PAS (prospective activity) occurs even earlier, during the cue presentation. This leads to strong correlations between visual responses to pair-associate stimuli.

Synaptic kinetics affects time scale of transitions in individual trials

To study the dependence of the dynamics of transitions on the kinetics of excitatory synapses, we varied the fraction of slow (100 ms, NMDA-like) to fast (2 ms, AMPA-like) excitatory recurrent currents, keeping constant the pair-learning parameter a. In Fig. 9A, one observes the rise of prospective population activity at various levels of x. As x increases, the 'prospective' activity rises more slowly, due to the slower dynamics of recurrent excitation, as is made clearer when the activity is synchronized at the transition time (Fig. 9B). The choice neurons take approximately 500 ms to complete the transition from the SAS to the elevated activity state, compared with 100 ms for a low fraction of slow recurrent excitation. As a consequence, when slow excitation is significant, the dynamics at the level of single trials becomes similar to the trial-averaged dynamics.

In conclusion, the basic mechanism of stochastic transitions between attractor states does not depend on the fraction of slow receptors, but the time course of the transition does. Prospective activity can rise sharply or gradually in the course of single trials. In both cases, the average of activity over many trials shows a ramping up of activity in neurons selective to the choice stimulus.

Individual spike trains distinguish fast/slow transitions

Can these two scenarios be distinguished experimentally? In experiment, spike trains of single neurons are recorded in single trials. A possible procedure is to select cells that show prospective activity. In each delay period of a trial in which the corresponding predictor stimulus is shown, the instantaneous spike rate is computed, using a bin size that should be longer than the average inter-spike interval, yet shorter than the average lifetime of the individual attractor states. Such a distribution of rates would be bi-modal for abrupt transitions. The peak at low rates would correspond to the time spent in the individual attractor state of the predictor, before the transition. The peak at high rates would correspond to the part of the delay interval following the transition. In the states reached after the transition, the choice neurons have elevated activity. On the other hand, slow transitions give rise to unimodal distribution of rates, due to the fact that in all trials the rate of choice neurons rises gradually from spontaneous activity to elevated persistent activity. In Fig. 10, we compare the histograms from network simulations in two cases, one with fast transitions ($x = 0.05$), the other with slow transitions ($x = 0.2$). Parameters are chosen such that the time course of trial-averaged prospective activity is similar in both cases (see Fig. 10A and B). Figure 10C is the single-cell rate histogram for low fraction of slow receptors (x), and hence a fast transition and the corresponding bi-modal distribution. Figure 10D is the histogram of the single-cell rate for the gradual case, which is unimodal.

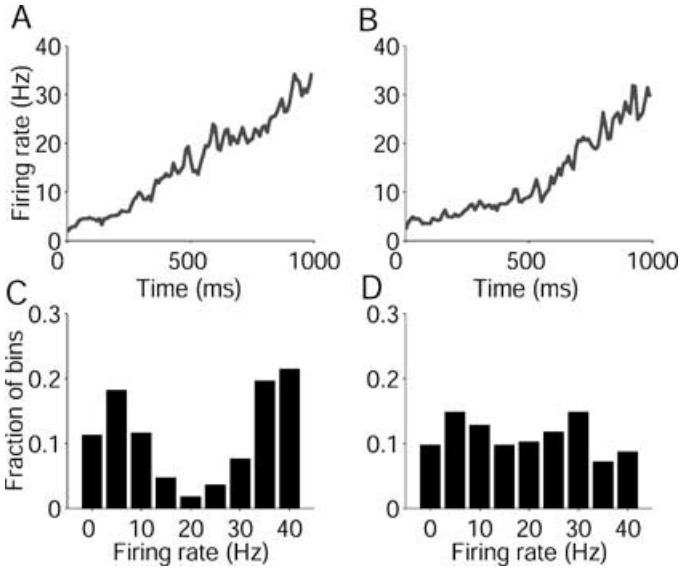

FIG. 10. Distinguishing fast/slow transitions in single-cell recordings. Trial-averaged prospective activity (A and B) and rate histograms for single cell during the delay period, in bins of 200 ms (C and D). (A and C) $x = 0.05$, $a = 0.15$. (B and D) $x = 0.2$, $a = 0.25$. Despite the similarity in the time course of prospective activity, in the case of fast transition ($x = 0.05$), single-cell rate distribution is bi-modal (A and C), indicating low rates at beginning of interval and high after a short transient. For slow transition ($x = 0.2$), the distribution is unimodal (B and D), as gradual rise samples all rates until saturation.

Symmetric vs. asymmetric synaptic structuring

The results exposed refer to simulations with prestructured synaptic matrix in the fully asymmetric condition, i.e. $a' = 0$. We also carried out simulations with a symmetric prestructured synaptic matrix. The main difference is that retrospective activity is more likely to survive when transitions occur in the delay period.

Learning in the Erickson–Desimone protocol

Next we perform a full simulation in which the neural dynamics is accompanied by synaptic plasticity, to mimick 2 days of the experiment of Erickson & Desimone (1999). In this process, we can monitor the evolution of the neural activities in different stages of the on-line learning process, which can be compared with experiment. We go further and monitor the evolution of the synaptic structure, beyond experimental access, to see if it actually converges to the type of structures assumed in the previous section; expose the different stages of the structuring; and check the asymptotic stability of the evolving synaptic structure.

The structuring is monitored by the fraction of potentiated synapses in the various homogeneous synaptic populations. Of interest are three types of populations of excitatory-to-excitatory synapses: synapses connecting neurons selective to the same stimulus (A → A); synapses connecting a neuron selective for a predictor stimulus to a neuron selective for a choice stimulus (A → A′); and, finally, synapses connecting neurons which are selective for stimuli which belong to different pairs (A → B). The fraction of potentiated synapses in these three populations of synapses are denoted by $C_{A\to A}$, $C_{A\to A'}$ and $C_{A\to B}$. In each trial, a small fraction of excitatory synapses switch from low to high state (LTP) or from high to low state (LTD), due to predictor and/or choice presentation.

In the following, we show the results obtained with fully asymmetric rule, i.e. $p_{w'} = 0$. Synaptic plasticity with the symmetric rule leads to results that are qualitatively rather similar.

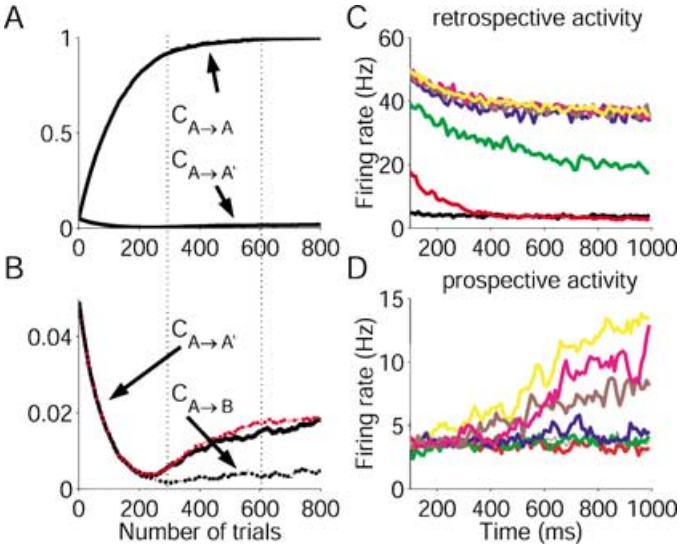

FIG. 11. Evolution of synaptic matrix and neuronal dynamics during learning. (A and B) Synaptic structuring vs. number of trials. (A) Fraction of potentiated intra-population synapses. $C_{A\to A}$ increases monotonically with the number of presentations until it reaches the saturated level $C_{A\to A}=1$. (B) Fraction of potentiated inter-population synapses for 100% (red) and 85% (black) of valid trials. $C_{A\to A'}$ decreases monotonically from the initial level $C_{A\to A'}=0.05$ until retrospective activity becomes stable, then increases with the number of presentations. It does not reach the level $C_{A\to A'}=1$, because also LTD takes place among inter-population connections. $C_{A\to A'}$ with 85% of valid trials is, on average, lower than that with 100% of valid trials, because of extra LTD induced by the non-valid trials. (C and D) Characteristics of neural activity vs. number of trials, for 85% of valid trials. (C) Retrospective activity average across successive 100 trials. Colour code: black 0–100; red 100–200; green 200–300; blue 300–400; brown 400–500; magenta 500–600; yellow beyond 600. Retrospective activity begins to appear between 200 and 300 trials (green curve), and becomes stable after about 400 trials. Further trials do not affect retrospective activity. (D) Prospective activity appears only after retrospective activity is in place, between trials 400 and 500 (brown curve). Network parameters as in Materials and methods, with $x=0.05$.

Three stages in the evolution of the network

In the first stage (trials 0–300, Fig. 11A and B), synaptic potentiation occurs only in synapses connecting neurons selective to the same stimulus, as all other neurons are at spontaneous levels (see Fig. 3, regions). The fraction of potentiated synapses in this synaptic population, $C_{A\to A}$ increases monotonically with the number of presentations, reaching saturation at $C_{A\to A}=1$, Fig. 11A. In the same period, LTD takes place in synapses connecting the neurons responsive to different images, and hence both $C_{A\to A'}$ and $C_{A\to B}$ decrease monotonically (Fig. 11B). In this phase, the presentation of a stimulus evokes only visual response at high rate in the corresponding population for the duration of the presentation. As soon as the stimulus is removed, emission rates decay back to the spontaneous level (Fig. 11C, curves black, red), as the average strength of synapses connecting neurons selective for the stimulus is not yet large enough to sustain delay activity. This stage continues until the resulting synaptic structure renders a state of persistent activity stable after the stimulus is removed, i.e. during the delay period. Retrospective delay activity appears.

In the second stage (trials 300–600, Fig. 11A and B), retrospective activity has become stable and allows synaptic potentiation in the inter-population connections. The enhanced emission rate of the predictor populations persists until the presentation of the choice stimulus (Fig. 11C, curves green, blue, brown and magenta). The neurons coding for the predictor are thus active at elevated rates, in close temporal proximity to those coding for the choice and synaptic potentiations can take place in (predictor → choice) synapses (see Materials and methods). The fraction of potentiated synapses $C_{A\to A'}$ begins to increase with the number of pairings, as shown in Fig. 11B. It does not reach saturation $C_{A\to A'}=1$, because LTD also takes place in this synaptic population, when the delay activity of the predictor has returned to spontaneous level while the choice population is still emitting at high rate. LTD also takes place in trials in which the predictor is not followed by the corresponding choice (invalid trials, see Materials and methods). This leads to an asymptotic fraction of potentiated synapses in the inter-population connections $C_{A\to A'}<<1$, whose value depends on the ratio between the probabilities of LTP and LTD and reflects the balance between potentiating and depressing processes in these synaptic populations. The asymptotic level of potentiated synapses in the inter-population connections corresponds to the forward pair-learning parameter a. When the fraction of synapses $C_{A\to A'}$ becomes of the order of 0.02, transitions between states become possible in the delay period. Prospective activity appears, as shown in Fig. 11D.

In the third stage (Fig. 11A and B, beyond trial no. 600), the network has reached an asymptotic synaptic structure. This structure may still fluctuate from trial to trial, due to random LTP and LTD transitions, especially in synapses connecting predictor to choice neurons, but the global variables $C_{A\to A}$, $C_{A\to A'}$ and $C_{A\to B}$ remain essentially constant. As in the case of the simulations with a fixed synaptic matrix, retrospective and/or prospective activity occur in individual trials.

The asymptotic level of potentiated synapses, $C_{A\to A'}$, depends on the percentages of pairings between the images during the training stage (valid trials). Figure 11B shows the evolution of $C_{A\to A'}$ with the number of trials, for two different percentages of valid trials in the protocol, 100% (red line) and 85% (black line). Lower percentage of valid pairings leads to lower percentage of potentiated synapses, $C_{A\to A'}$.

Statistical analysis of spike rates

To compare the evolution of the neural activity patterns in the course of training in the simulation with Erickson & Desimone (1999), we use the average correlation between visual responses to pair-associate stimuli; between predictor visual response and delay activity and between choice visual response and delay activity, vs. the number of trials, in the course of training. We estimate the average rates of a sample (10%) of cells in predictor, delay and choice periods, separately for each of the eight (predictor–choice) pairs. The sample contains the same number of cells for each pair. For each cell and in each trial, the rate during cue and test presentations is estimated in a window 75 ms to 250 ms from the presentation; delay period rate is estimated in a window 200 ms after cue removal to 200 ms before test presentation. The average responses are obtained by averaging single-cell rates across trials with the same pair of stimuli, for each cell.

The simulation is divided into successive groups of 100 trials, and correlations between predictor and choice rates, predictor and delay rates, and delay and choice rates are computed in each group. The correlation between the predictor visual response and the delay activity begins to increase right from the beginning (Fig. 12), while both correlation between visual responses and between delay activity and choice visual response remain initially at chance level. After the first 300 trials, delay activity is significantly correlated with the response to the predictor stimulus (Fig. 12). This is due to the presence of retrospective activity (see also Fig. 11). As the training proceeds, the correlation between the visual responses to paired stimuli increases (Fig. 12). Similarly, the correlation between choice visual response and

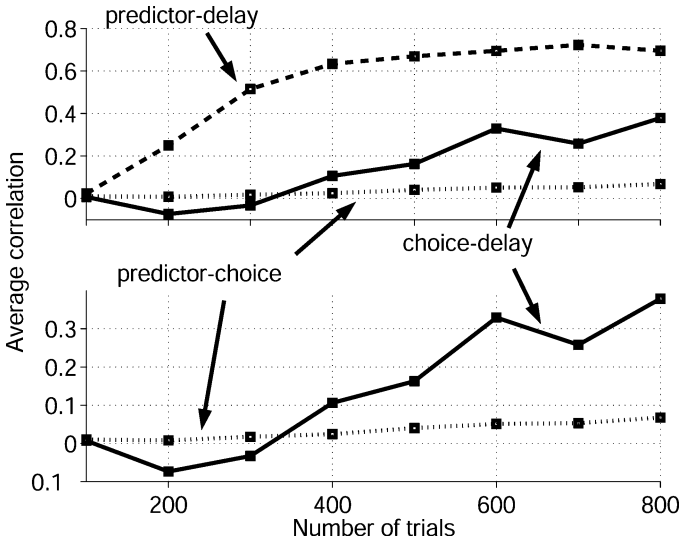

FIG. 12. Average correlations between visual responses to paired stimuli, between predictor response and delay activity and between choice response and delay activity, vs. number of trials. Predictor response and delay activity become strongly correlated as soon as retrospective persistent activity becomes stable, i.e. around trials 200–300. It then reaches an asymptotic high level of about 0.7. Correlations between predictor and choice visual responses and between choice response and delay activity begin to increase between 400 and 500 trials, only after correlation between predictor response and delay activity reaches a high level, i.e. retrospective activity has appeared. The correlation between choice response and delay activity reaches an asymptotic level of about 0.3, while the correlation between predictor and choice visual responses reaches about 0.07, only after trial 500. Average values are computed in each successive group of 100 trials separately. See text for more details.

the delay activity significantly increases (Fig. 12). In contrast, the correlation between predictor visual response and the delay activity reaches a steady level, which is not substantially affected beyond 400 trials.

These results are a direct consequence of the increase of the pair-learning parameter a in the course of training. As can be seen from Fig. 11, after about 200–300 trials, the ($A \rightarrow A$) connections reach a potentiation level that sustain retrospective activity. At the same time, the ($A \rightarrow A'$) connections remain in their low state. As training proceeds, the mean level of potentiation $C_{A \rightarrow A'}$ increases, reaching the asymptotic level after 500–600 trials (corresponding to a in the prestructured synaptic matrix), leading to prospective activity.

Discussion

In the present paper, we have investigated the learning dynamics of a cortical network model subjected to the pair-associate protocol. In a first stage, the plasticity leads to the formation of neural representations for single images (selective delay activity). During the visual presentation, the concurrent activation at high rate of the cells coding for the same stimulus causes potentiation of the synapses connecting the neurons activated by the same stimulus. When the efficacy within each of these synaptic populations reaches a suitably high level, the neural population becomes capable of sustaining reverberating activity, in the absence of external inputs. This persistent activity maintains an active memory of a stimulus shown in the past (the 'predictor' stimulus) – retrospective activity. These states are attractor states, expressing the fact that a large variety of neighbouring stimuli evoke the same self-sustained distribution of the level of average rates.

Once retrospective activity becomes stable, it persists across the delay interval, until the presentation of the choice stimulus. The delay period activity allows neurons coding for the predictor to be active in close temporal proximity to the visual response of the choice neurons, which leads to the potentiation of the synapses between the two neural populations (inter-population connections). This potentiation is weak, relative to that within each population, and its amplitude is governed by the percentage of trials in which the second image is the fixed pair-associate (in contrast to those where it is a randomly chosen image). It is important to point out that the level of inter-population potentiation reaches a stable (low) asymptotic level during learning, governed by a balance between LTP and LTD. LTD in the inter-population synapses intervenes either during the cue presentation, or in the part of the choice presentation when the retrospective delay activity had died down, or in the course of non-valid trials (see Results). None of these bring about LTD in intra-population synapses. These connections give rise to transitions, after the cue presentation, to other types of persistent states available to the network: either pair states, combining the neurons of both the predictor and the choice of the corresponding pair, or the individual persistent activity state of the choice stimulus. The activation of the choice neurons, prior to the presentation of the choice stimulus, is referred to as prospective activity (see, e.g. Fuster, 2001 and refs therein). In the first scenario retrospective activity persists, while in the second it dies out during the delay period.

The transitions are caused by the fluctuations in the neural spiking dynamics. The probability of occurrence of such a transition depends mainly on the strength of inter-population connections and on the level of noise in the system. As the strength of inter-population connections increases, the basin of attraction of the IAS shrinks in favour of that of the corresponding pair states, rendering the transitions more and more frequent. As a matter of fact, stationary pair states exist even without inter-population strengthening (Amit *et al*., 2003). Yet for transitions due to noise, from an individual state to the associated-pair state, to occur selectively and with high likelihood within a 1-s delay period, a sufficiently high level of inter-population potentiation is required.

In this account, the potentiation of the inter-population connections depends on the existence of the predictor persistent state. The appearance of the retrospective activity, prior both to the pair-coding neurons and to the prospective activity, is therefore a logical prerequisite and prediction (see below) of the scenario proposed. In the absence of the predictor persistent state, no inter-population potentiation is possible.

In networks of neurons spiking asynchronously, the fluctuations in population activity are due to the finite size of the network and would vanish were the network to become very large. However, these fluctuations can be large, even for networks of a size similar to cortical modules (Brunel & Hakim, 1999; Brunel, 2000a).

Comparison with experiment

Prospective activity in PRh cortex and IT

Our model reproduces most of the available neuro-physiological data obtained during delayed pair-associate tasks in the temporal lobe of the monkey. Sakai & Miyashita (1991) and Naya *et al*. (2001, 2003) found two types of cells in their recordings of area TE of IT cortex and of area 36 of PRh cortex: pair-coding cells and pair-recall cells. Our model explains the response characteristics of both types of cells in a unifying framework.

Both types of cells arise due to learning dynamics, which potentiate the connections between cells that are selective for pair-associate stimuli, while the relative occurrence of both types is related to the magnitude of the pair-learning parameter. If a is small (of order 0.02 in the simulations of Fig. 8), transitions from the predictor attractor to a

pair-associate attractor (either individual or pair state) take place only during the delay period. The visual responses to paired stimuli are only weakly correlated, while delay activity is strongly correlated with choice response. Such neurons have all the characteristics of pair-recall neurons.

By contrast, if the pair-learning parameter is large (of order 0.04 in the simulations of Fig. 8), transitions to the 'prospective' states occur during the cue presentation, yielding strong correlations between the visual response to pair-associate stimuli. Hence, at large a, cell activities have the characteristics of pair-coding neurons.

Naya *et al*. (2001) found that neurons in PRh cortex are typically pair-coding neurons (see also Naya *et al*., 2003), while neurons in area TE are typically pair-recall neurons. Differences between these areas could be due to differences in learning dynamics between areas 36 and TE. Naya *et al*. (2001) hypothesize that backward projections from PRh cortex to IT are responsible for the transitions during the delay period to prospective activity in IT. Lesions in PRh cortex suppress correlations between visual responses to pair-associates in IT (Higuchi & Miyashita, 1996), lending further support to the role of backward projections.

We would argue that these data could be accounted for if one assumes that in area TE the pair-learning parameter is small (or even zero). In that case, in the absence of backward projections from PRh cortex, the average transition times are very long. Backward projections would then provide a biased input favouring transitions to prospective activity in area TE. In the present model, the gradual rise of activity seen in pair-recall neurons (Sakai & Miyashita, 1991) is compatible with two scenarios: one in which the rise is gradual on a single trial basis, and another in which the rise is abrupt in a single trial, but occurs at random instances during the delay, and thus is gradual when averaged across trials (see Fig. 9). Naya *et al*. (2001) provide evidence that the transition duration is short, of order 100 ms. This would indicate that in area TE, the fraction of slow excitatory receptors is small.

The studies of Erickson & Desimone (1999) and Messinger *et al*. (2001) provide evidence that modifications of neuronal selectivity due to learning of new associations can occur on the time scale of hours. Our model operates at the same time scale of appearance of retrospective and prospective activity seen in the experiment. In the model, this time scale is related to LTP and LTD transition probabilities. Our results are consistent in more detail. During the learning of a new pair, there is a first stage (first day) in which the delay activity in the PRh cortex is correlated only with the predictor ('retrospective' working memory), while after a second day of training, delay activity is correlated also with the choice ('prospective' activity). Our model accounts for these two distinct stages: indeed, the presence of 'retrospective' activity is a prerequisite, before synapses connecting populations of cells selective for pair-associates can potentiate, and hence 'prospective' activity starts to develop. The simulation experiments reported in Fig. 11 show that during the first 400 trials (corresponding to the 'novel' condition, 1 day in the experiment of Erickson & Desimone, 1999), 'retrospective' activity has already become robust, while 'prospective' activity has barely appeared. In the next 400 trials (corresponding to 'familiar' condition, second day of the experiment), 'prospective' activity becomes prominent, as transitions between IAS and PAS occur quite often during the delay period.

Our model also accounts for structure of correlations in spike rates. In the 'novel' condition, i.e. following a relatively short training, the delay activity is correlated only with the predictor visual response. By contrast, in the 'familiar' condition, i.e. following a relatively long training (2 days), the delay activity becomes correlated also with the choice visual response. The magnitudes of the correlations in our simulations are rather similar to the experimentally observed ones: the average correlation between predictor and delay is in experiment 0.316 for 'novel' and 0.404 for familiar stimuli (0.36 average over all first 400 trials, and 0.69 average over the last 400 trials, in our simulations); the correlation between choice visual response and delay activity is 0.079 for 'novel', 0.269 for 'familiar' (0.002 and 0.28, in our simulations); the correlations between the visual responses of predictor and choice are −0.002 for 'novel' and 0.145 for 'familiar' stimuli (0.015 and 0.053 in our simulations). In accordance with the experiment, the correlation between visual responses does not account for the correlation between choice visual response and delay activity. This is due to the fact that at this relatively low level of a, visual responses to pair-associates are weak, and most 'prospective' effects occur during the delay period and not during the cue period.

There are two significant quantitative differences between simulations and experiment (i) Between predictor visual response and delay, for familiar stimuli (0.404 vs. 0.69). This difference could be due to the fact that in the experiment retrospective activity dies out more often. It would be remedied by a somewhat lower value of excitatory potentiation or a higher value of the pair-association parameter. (ii) Between 'predictor' and 'choice' visual responses for familiar stimuli (0.145 vs. 0.053). This difference could be explained by differences in the magnitude of the rate of selective visual responses. In our simulations, the correlations between visual responses increase from 0.053 to 0.119 if the visual response of selective neurons is decreased from 160 Hz to 80 Hz.

Some studies have failed to find evidence of associative learning, i.e. cells exhibiting 'prospective' activity. Gochin *et al*. (1994) used a protocol similar to that of Sakai & Miyashita (1991) and Erickson & Desimone (1999), with the difference that individual stimuli were used in more than one pair. Our model would account for the absence of 'prospective' activity, as if the percentage of trials in which the two stimuli are paired is lowered, the pair-learning parameter does not reach the threshold to produce significant prospective activity. Another possible reason for discrepancies between different studies stems from differences between different areas of the temporal lobe, and in particular between PRh cortex (area 36) and area TE of IT cortex (Naya *et al*., 2001, 2003).

Prospective activity in PFC

PFC has long been involved in the expectation and preparation of anticipated events (see, e.g. Fuster, 2001 and refs therein). Prospective activity, i.e. increased firing of cells in apparent anticipation of the motor response or another stimulus related to it, has been observed in PFC (Niki & Watanabe, 1979; Fuster *et al*., 1982, 2000; Sawaguchi *et al*., 1989; Quintana & Fuster, 1999; Rainer *et al*., 1999). Changes of neuronal activity in the delay period have been shown to arise due to associative learning (Asaad *et al*., 1998; Rainer *et al*., 1999; Fuster *et al*., 2000). As training takes place, the delay activity shifts from purely retrospective to prospective, and the shift takes place dynamically during the delay period (Rainer *et al*., 1999). This is again consistent with our findings. Indeed, in our model, activity is mostly 'retrospective' until the beginning of the delay period, and becomes more and more 'prospective' as the network has time to make transitions to the PAS, or to gradually move into these states when recurrent excitatory synapses have a sufficiently strong slow component. These data suggest that the basic mechanisms of the generation of prospective activity, through the inter-play between retrospective persistent activity and Hebbian learning, apply also to PFC. This, despite significant functional differences between PFC and areas of the temporal lobe, such as the facts that in PFC, cells represent not only the external stimuli but also motor responses and errors, and cell responses are less selective.

Experimental predictions of the model

The main prediction from our study is that in delayed response tasks, prospective activity can only appear if retrospective activity is stable. This prediction could be tested experimentally by manipulating persistent activity in the delay period, using iontophoresis of any neurotransmitter that is known to affect persistent activity, such as dopamine (Williams & Goldman-Rakic, 1995) or GABA (Rao *et al*., 2000). Iontophoresis, leading to suppression of persistent activity, should be applied in the 'first day' with 'novel' stimuli, in the experimental protocol of Erickson & Desimone (1999). In the second day, no iontophoresis should be performed. The prediction is that in the second day, retrospective activity would be observed, but no prospective activity. The appearance of prospective activity would be delayed by the blockade of persistent activity due to iontophoresis.

Our study makes a second clear prediction: as the pair-learning parameter increases, the correlation between visual responses to paired stimuli should also increase, while the time of appearance of prospective activity (measured from cue onset) should decrease. The magnitude of a can be manipulated by varying the relative frequency of trials in which pair-associate stimuli are shown together ('valid trials', in Erickson & Desimone, 1999). As the percentage of 'valid trial' increases so should the mean correlation between visual responses to pair-associate images. Correspondingly, ′prospective′ activity should appear earlier in a trial or, equivalently, the slope of the rise of averaged prospective activity should increase, as shown in Fig. 8.

A third prediction is that if fast excitatory synaptic transmission predominates, the transition in a given trial should be very steep, and not gradual as seen on average. An alternative scenario would be that the increase of activity is gradual in every trial, as is the case for higher proportions of slow excitatory synaptic component. The type of transition can be identified in experiments with single-cell recording by analysing the binned spike rate histogram of a single neuron over several trials (Fig. 10). A similar procedure was used by Chafee & Goldman-Rakic (1998) to characterize a slow increase of persistent activity in the delay period of a delayed oculomotor task. Manipulation of NMDA and/or AMPA levels in parallel with neurophysiological recordings during pair-associate tasks may put this prediction to a test.

One would also expect that, if 'prospective' activity is actually related to behaviour, part of the variability in the reaction time, as well as in the performance level, of the monkey could be due to the variability in the transition times in the delay period. This should be true particularly in the early stage of learning, when a is expected to be low. Our study predicts that for low a, the transition to the pair-associate state occurs during the delay period, with a probability that depends on a. Thus, one can expect shorter reaction times and higher performance level, correlated with instances when the transition actually took place in the delay period, with respect to the instances when transitions did not take place.

Theoretical issues

Synaptic plasticity

A serious limitation of the present study is that the synaptic plasticity mechanism is still rudimentary, as what drives synaptic changes at individual synapses is average rates (in sliding windows of 100 ms) of pre- and post-synaptic neurons. Recently, much experimental work has been devoted to the details of what actually controls synaptic changes, at the level of pre- and post-synaptic spike trains (Markram *et al*., 1997; Bi & Poo, 1998; Sjöström *et al*., 2001). Thus, one may expect that in the near future plasticity will be better grounded in the biophysics and biochemistry of synapses. Such mechanisms should then be incorporated in network studies to confirm that the dynamics of persistent activity as shown in this paper is indeed a realistic scenario. An indication that this is feasible is provided by a study showing that spike-driven synaptic dynamics (Fusi *et al*., 2000) succeed in generating a synaptic structure that sustains retrospective activity (Amit & Mongillo, 2003).

Non-overlapping vs. overlapping stimuli

We have used non-overlapping stimuli, in the sense that a neuron responds visually to at most one stimulus. This choice is made for several reasons, beside its simplicity: (i) in the temporal lobe, cells that display delay activity are typically very selective, often for only one of the stimuli involved; (ii) preliminary simulations with randomly overlapping stimuli show that the formation of retrospective activity occurs as in the case of non-overlapping stimuli (Mongillo and Amit, unpublished results).

Transitions during delay period

On the theoretical side, most studies of persistent activity have no dynamical effects in the delay period, but focused on the properties of stationary attractor states. Recently, Koulakov (2001) has studied the degradation of delay activity due to the unreliability of synaptic transmission. Reutimann *et al*. (2001) have interpreted the rise of spike rates in some cells during delay period in a experiment (the 'expectation cells') as due to short-term synaptic dynamics during the delay period. Noise-driven transitions between selective attractor states had only been previously considered in networks with binary neurons (Buhmann & Schulten, 1987; Amit, 1988). In networks with continuous rather than discrete attractors, random drifts of the network state are observed in the presence of noise (Ben-Yishai *et al*., 1995; Seung, 1996; Camperi & Wang, 1998; Compte *et al*., 2000; Laing & Chow, 2001), due to the translational invariance of the continuous attractor.

The present work shows a richer and more dynamical picture of persistent activity. Previously, delay period activity was considered a fast relaxation towards a fixed-point attractor, used as a vehicle for working memory. Including Hebbian learning and allowing for transitions between attractor states, which are nearby in state space, changes significantly the picture. The system explores the space during the delay period, as a consequence of fluctuations. Transitions are not made to arbitrary attractors (which would be a rather pathological situation for a memory system), but rather to states that have been linked by associative Hebbian learning. These transitions may form the substrate of cognitive operations used when stimulus–stimulus associations are required. Learning allows the system to 'garden' its attractor landscape, allowing barriers between attractors representing associated stimuli to be lowered, and hence transitions between these states become easier. As a result, the system becomes capable of predicting the appearance of future stimuli on the basis of past experience.

Acknowledgements

This study was supported by the Center of Excellence Grant 'Changing Your Mind' of the Israel Science Foundation and the Center of Excellence Grant 'Statistical Mechanics and Complexity' of the INFM, Roma-1. We thank Emanuele Curti for help in the simulations, Alberto Bernacchia and Massimo Mascaro for useful comments on an early version of the manuscript. N.B. wishes to thank Stefano Fusi and Xiao-Jing Wang for helpful discussions.

Abbreviations

GABA, γ-aminobutyric acid; IAS, individual attractor states; IF, integrate-and-fire; IT, ; LTP, long-term potentiation; NMDA, *N*-methyl-D-aspartate; PAS, pair attractor states; PFC, prefrontal cortex; PRh, perirhinal; SAS, spontaneous activity state.

References

Amit, D.J. (1988) Neural networks counting chimes. *Proc. Natl Acad. Sci.*, **85**, 2141–2145.

Amit, D.J. (1995) The hebbian paradigm reintegrated: local reverberations as internal representations. *Behav. Brain Sci.*, **18**, 617.

Amit, D.J. (1998) Simulation in neurobiology: theory or experiment? *Trends Neurosci.*, **21**, 231–237.

Amit, D.J., Bernacchia, A. & Yakovlev, V. (2003) Multiple-object working memory – a model for behavioral performance. *Cerebral Cortex*, **13**, 435–443.

Amit, D.J. & Brunel, N. (1997a) Dynamics of a recurrent network of spiking neurons before and following learning. *Network*, **8**, 373–404.

Amit, D.J. & Brunel, N. (1997b) Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, **7**, 237–252.

Amit, D.J., Brunel, N. & Tsodyks, M.V. (1994) Correlations of cortical hebbian reverberations: experiment vs theory. *J. Neurosci.*, **14**, 6435–6445.

Amit, D.J. & Mongillo, G. (2003) Spike driven synaptic plasticity generating working memory states. *Neural Computation*, **15**, 565–596.

Asaad, W.F., Rainer, G. & Miller, E.K. (1998) Neural activity in the primate prefrontal cortex during associative learning. *Neuron*, **21**, 1399–1407.

Ben-Yishai, R., Lev Bar-Or, R. & Sompolinsky, H. (1995) Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. USA*, **92**, 3844–3848.

Bi, G.-Q. & Poo, M.-M. (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.*, **18**, 10,464–10,472.

Brunel, N. (1996) Hebbian learning of context in recurrent neural networks. *Neural Computation*, **8**, 1677–1710.

Brunel, N. (2000a) Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.*, **8**, 183–208.

Brunel, N. (2000b) Persistent activity and the single cell {f-I} curve in a cortical network model. *Network*, **11**, 261–280.

Brunel, N. & Hakim, V. (1999) Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation*, **11**, 1621–1671.

Buhmann, J. & Schulten, K. (1987) Noise driven temporal association in neural networks. *Europhysics Letters*, **4**, 1205.

Camperi, M. & Wang, X.-J. (1998) A model of visuospatial short-term memory in prefrontal cortex: recurrent network and cellular bistability. *J. Comput. Neurosci.*, **5**, 383–405.

Chafee, M.V. & Goldman-Rakic, P.S. (1998) Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J. Neurophysiol.*, **79**, 2919–2940.

Compte, A., Brunel, N., 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.

Durstewitz, D., Seamans, J.K. & Sejnowski, T.J. (2000) Neurocomputational models of working memory. *Nature Neurosci. Suppl.* 1184–1191.

Erickson, C.A. & Desimone, R. (1999) Responses of macaque perirhinal neurons during and after visual stimulus association learning. *J. Neurosci.*, **19**, 10,404–10,416.

Ermentrout, G.B. (1992) Complex dynamics in winner-take-all neural nets with slow inhibition. *Neural Networks*, **5**, 415–431.

Ermentrout, G.B. (1996) Type I membranes, phase resetting curves, and synchrony. *Neural Computation*, **8**, 979–1001.

Ermentrout, G.B. (1998) Neural networks as spatio-temporal pattern-forming systems. *Report Prog. Phys.*, **61**, 353–430.

Funahashi, S., Bruce, C.J. & Goldman-Rakic, P.S. (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.*, **61**, 331–349.

Fusi, S., Annunziato, M., Badoni, D., Salamon, A. & Amit, D.J. (2000) Spike-driven synaptic plasticity: theory, simulation, VLSI implementation. *Neural Computation*, **12**, 2227–2258.

Fuster, J.M. (1995) *Memory in the Cerebral Cortex*. MIT Press, Cambridge, MA.

Fuster, J.M. (2001) The prefrontal cortex – an update: time is of the essence. *Neuron*, **30**, 319–333.

Fuster, J.M. & Alexander, G. (1971) Neuron activity related to short-term memory. *Science*, **173**, 652–654.

Fuster, J.M., Bauer, R.H. & Jervey, J.P. (1982) Cellular discharge in the dorsolateral prefrontal cortex of the monkey in cognitive tasks. *Exp. Neurol.*, **77**, 679–694.

Fuster, J.M., Bodner, M. & Kroger, J.K. (2000) Cross-modal and cross–temporal association in neurons of frontal cortex. *Nature*, **405**, 347–351.

Fuster, J.M. & Jervey, J.P. (1981) Inferotemporal neurons distinguish and retain behaviourally relevant features of visual stimuli. *Science*, **212**, 952–955.

Gochin, P.M., Colombo, M., Dorfman, G.A., Gerstein, G.L. & Gross, C.G. (1994) Neural ensemble coding in inferior temporal cortex. *J. Neurophysiol.*, **71**, 2325–2337.

Goldman-Rakic, P.S. (1995) Cellular basis of working memory. *Neuron*, **14**, 477–485.

Griniasty, M., Tsodyks, M.V. & Amit, D.J. (1993) Conversion of temporal correlations between stimuli to spatial correlations between attractors. *Neural Computation*, **5**, 1–17.

Hansel, D. & van Vreeswijk, C. (2002) How noise contributes to contrast invariance of orientation tuning in cat visual cortex. *J. Neurosci.*, **22**, 5118–5128.

Hebb, D.O. (1949) *Organization of Behavior.* Wiley, New York.

Higuchi, S. & Miyashita, Y. (1996) Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proc. Natl Acad. Sci. USA*, **93**, 739–743.

Koulakov, A. (2001) Properties of synaptic transmission and the global stability of delayed activity states. *Network*, **12**, 47–74.

Laing, C.R. & Chow, C.C. (2001) Stationary bumps in networks of spiking neurons. *Neural Computation*, **13**, 1473–1494.

Markram, H., Lubke, J., Frotscher, M., Roth, A. & Sakmann, B. (1997) Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol. (Lond.)*, **500**, 409–440.

Mason, A., Nicoll, A. & Stratford, K. (1991) Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. *J. Neurosci.*, **11**, 72–84.

McCormick, D., Connors, B., Lighthall, J. & Prince, D. (1985) Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons in the neocortex. *J. Neurophysiol.*, **54**, 782–806.

Messinger, A., Squire, L.R., Zola, S.M. & Albright, T.D. (2001) Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proc. Natl Acad. Sci. USA*, **98**, 12239–12244.

Miyashita, Y. (1988) Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, **335**, 817–820.

Miyashita, Y. & Chang, H.S. (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, **331**, 68–70.

Nakamura, K. & Kubota, K. (1995) Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. *J. Neurophysiol.*, **74**, 162–178.

Naya, Y., Sakai, K. & Miyashita, Y. (1996) Activity of primate inferotemporal neurons related to a sought target in pair association task. *Proc. Natl Acad. Sci. USA*, **93**, 2664–2669.

Naya, Y., Yoshida, M. & Miyashita, Y. (2001) Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science*, **291**, 661–664.

Naya, Y., Yoshida, M. & Miyashita, Y. (2003) Forward processing of long-term associative memory in monkey inferotemporal cortex. *J. Neurosci.*, **23**, 2861–2871.

Niki, H. & Watanabe, M. (1979) Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Res.*, **171**, 213–224.

Quintana, J. & Fuster, J.M. (1999) From perception to action: temporal integrative functions of prefrontal and parietal neurons. *Cereb. Cortex*, **9**, 213–221.

Rainer, G., Rao, S.C. & Miller, E.K. (1999) Prospective coding for objects in primate prefrontal cortex. *J. Neurosci.*, **19**, 5493–5505.

Rao, S.G., Williams, G.V. & Goldman-Rakic, P.S. (2000) Destruction and creation of spatial tuning by disinhibition: $GABA_A$ blockade of prefrontal cortical neurons engaged by working memory. *J. Neurosci.*, **20**, 485–494.

Reutimann, J., Fusi, S., Senn, W., Yakovlev, V. & Zohary, E. (2001) A model of expectation effects in inferior temporal cortex. *Neurocomputing*, **38–40**, 1533–1540.

Sakai, K. & Miyashita, Y. (1991) Neural organization for the long-term memory of paired associates. *Nature*, **354**, 152–155.

Sawaguchi, T., Matsumura, M. & Kubota, K. (1989) Depth distribution of neuronal activity related to a visual reaction time task in the monkey prefrontal cortex. *J. Neurophysiol.*, **61**, 435–446.

Seung, H.S. (1996) How the brain keeps the eyes still. *Proc. Natl Acad. Sci. USA*, **93**, 13,339–13,344.

Sjöström, P.J., Turrigiano, G.G. & Nelson, S. (2001) Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, **32**, 1149–1164.

Wang, X.-J. (2001) Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.*, **24**, 455–463.

Williams, G.V. & Goldman-Rakic, P.S. (1995) Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature*, **376**, 572–575.

Wilson, H.R. & Cowan, J.D. (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, **12**, 1–24.