Cortical dynamics of memory

Joaquin M. Fuster*

UCLA Neuropsychiatric Institute, 760 Westwood Plaza, Los Angeles, CA 90024, USA

Received 23 March 1999; accepted 23 March 1999

Abstract

Memory networks are formed in the cerebral cortex by associative processes, following Hebbian principles of synaptic modulation. Sensory and motor memory networks are made of elementary representations in cell assemblies of primary sensory and motor cortex (phyletic memory). Higher-order individual memories, e.g. episodic, semantic, conceptual — are represented in hierarchically organized neuronal networks of the cortex of association. Perceptual memories are organized in posterior (post-rolandic) cortex, motor (executive) memories in cortex of the frontal lobe. Memory networks overlap and interact profusely with one another, such that a cellular assembly can be part of many memories or networks. Working memory essentially consists in the temporary activation of a memory network, as needed for the execution of successive acts in a temporal structure of behavior. That activation of the network is maintained by recurrent excitation through reentrant circuits. The recurrent reentry may occur within local circuits as well as between separate cortical areas. In either case, recurrence binds together the associated components of the network and thus of the memory it represents. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Memory; Cell assemblies; Episodic; Semantic and conceptual memories; Working memory; Memory networks

In this article I discuss the neural mechanisms by which cortical memory networks are activated in the retention of active memories at the service of behavior, in other words, in so-called ‘working memory.’ Two related subjects will be successively considered:

1. In the first place, by way of background, I will deal briefly with the architecture of memory networks, that is, with the static or anatomical aspects of representation of memory in the cerebral cortex.

2. Secondly, I will deal with the main topic of the article, the dynamics of those memory networks in active short-term, or working, memory.

1. Network memory

As we gather empirical evidence from human neuropsychology, from behavioral electrophysi-
ology in primates, and from the neuroimaging of active memory in the human, the concept of distributed cortical memory is becoming every day more compelling. To be sure, this is not an altogether new concept. It has been favored in the past by a number of experimentalists and theoreticians of the brain (Lashley, 1950; Hayek, 1952; Mountcastle, 1978). What is new is the increasing recognition that the memory networks are widely spread over the cortex, and that their distribution transcends any given cortical areas or module, however defined. Further, we are learning that those networks overlap one another anatomically, are highly interactive, and hierarchically organized. Most radically new is the idea that different networks share some of the same neurons and connections; in other words, that any cortical neuron or group of neurons can be part of many networks, and thus part of many memories.

First, let us recapitulate our best understanding of how the memory networks of the cortex are formed, how is memory acquired. At the root of memory formation is obviously the synapse. This was first hypothesized by Tanzi and by Cajal, in 1893–1894 and has now been well substantiated. A memory is encoded in the cortex by a pattern of interconnected neurons, and thus the conductivity of the synapses that web a network of neurons together is the essence of mnemonic encoding, of the memory trace. A memory is a network, which is structurally reducible to axons, cells and molecules, but the content of the memory is not reducible to those neural elements or even to their patterns of change or of cell firing. The content of the memory is a relational code that emerges from the combination of those elements and is essentially irreducible to them.

In the course of ontogenetic development, and also probably in phylogeny, the memory networks of primary sensory and motor systems of the cortex become established according to certain principles of selective reorganization. Those networks will constitute what I have called phyletic memory, which is the essential substrate of sensory and motor cortex and constitutes the foundation of the hierarchy of the memory networks of the individual. Above that phyletic base, in parasensory and paramotor areas, and in cortex of association, the networks of individual experience will be formed. These will be constituted by hebbian principles of synaptic potentiation and by self-organization in interaction with the environment.

We have good reasons to suspect, although it has not been proven, that in the formation of neocortical networks of individual memory, the hippocampus, a region of ancient cortex, plays a fundamental role (Amaral, 1987; Squire and Zola-Morgan, 1988). It is still a matter of speculation that this role is mediated by certain mechanisms such as LTP and certain excitatory, glutaminergic receptors, such as NMDA receptors.

The aggregate result of the primate’s interactions with the environment will be the gradual emergence, on a basis of phyletic memory, of two massive hierarchies of overlapping and profusely integrated networks: one hierarchy in posterior (postrolandic) neocortex for perceptual memory and the other in anterior or frontal neocortex for motor or executive memory. It is a curious and probably more than casual coincidence that the two hierarchies develop along myelogenetic gradients.

At the base of the connective hierarchy of perceptual memories lies a layer of phyletic sensory memory, the primary sensory cortices. Above that, in parasensory cortices, lies a layer of cross-modal and polysensory memories or networks. Higher up, in cortex of association, lie the networks of episodic, semantic, and conceptual memory, by order of increasing generality and abstraction.

‘Long-term’ memory, therefore, constitutes the sum-aggregate of all the connectivity established in those two massive, overlapping, interconnected and hierarchically organized networks of neocortical neurons. The idiosyncrasy, indeed individuality, of our memories derives from the combinatorial power, practically infinite, of some ten billion neocortical neurons. To be sure, on every new experience that combinatorial power is constrained by such factors as the organization and strength of prior synaptic connectivity.

Before continuing, I want to emphasize two important caveats that derive from the anatomy
of memory and that bear on the processing of memory: (a) hierarchical organization does not necessarily imply serial processing, for serial as well as parallel processing take place in the processing of any memory, and (b) the processing of perceptual and motor memories involves by necessity the cortico–cortical connections between posterior and anterior neocortex at all levels of the perceptual and executive hierarchies, in other words, the associative connections between perceptual and executive networks.

2. Active memory

In the course of our daily life, myriad networks of latent and passive long-term memory are retrieved and activated in our cortex by association, as required by perturbations in the internal and external milieu. Some are only transiently activated in conscious or unconscious recall or recognition. Others, because they are essential to the temporal structuring of reasoning and behavior, are activated in a sustained fashion, as required for the integration of acts and behaviors that depend on temporally discontinuous events. Such is the case in sequential behaviors, in the mental search for solution to logical problems and, of course, in the performance of that broad category of tasks invented by experimental psychologists that we commonly call `delay tasks.' These tasks have turned out to be invaluable for exploring the neuronal dynamics of active memory.

Clearly we are not quite ready yet to explore in depth the neural dynamics of episodic or semantic memory, let alone conceptual memory. Neuropsychology and neuroimaging merely provide us with faint glimpses of the mechanisms of their activation in the encoding and retrieval of memories. What we are ready to explore, and are exploring, is the mechanisms of activation of sensory memories in so-called working memory, and therefore in delay tasks. This kind of research, especially with microelectrodes, is not only useful for us to be able to study memory mechanisms but to unravel the topography and architecture of memory networks. The mechanisms of activation of sensory memories, we hope, are paradigmatic of the mechanisms of activation of other forms of memory, episodic and semantic for example.

It is in the prefrontal cortex where many years ago, in my laboratory, we found the first working-memory cells. However some misconceptions should be dispelled in this regard. The prefrontal cortex is not the center of working memory, working memory is not the only function of the prefrontal cortex, and working memory is not circumscribed to the prefrontal cortex. The reason the prefrontal cortex is so prominent in the working-memory literature is not only because memory cells were found there first, but because all the tasks used to test working memory in humans or monkeys (e.g. delay tasks) require the integration of a motor act with prior sensory information (Fuster, 1995). It is the temporal integration of the act, an executive action, that makes the prefrontal cortex and its cells so important for the working memory that the prospective execution of the act requires (Fuster, 1985).

Working memory is, in fact, a state of memory as widely distributed over the cortical surface as the long-term memory network that constitutes its structural substrate. Thus, for example in the monkey that has been trained to perform a visual delayed matching-to-sample task, the cells of inferotemporal cortex are activated and remain activated while the monkey must retain the memory of the visual stimulus (Fuster and Jervey, 1982). In addition, of course, prefrontal cells are also activated and remain activated because that stimulus elicits a motor memory and must be retained and activated, together with the motor memory, until the action is executed (Fuster et al., 1982). The breadth of the cortical network activated in visual short-term memory can also be observed by neuroimaging in the human (Swartz et al., 1995).

The shared activation by inferotemporal and prefrontal cells in visual delay matching reflects the interactions within a vast memory network of posterior and frontal cortex in visual working memory. It is thus reasonable to postulate that, in a visual delayed matching task, the memorandum, which is a stimulus to be retained for prospective action, elicits the activation of a large network with two component networks: an inferotemporal
network representing the visual memorandum and a prefrontal network representing the motor response at the end of each trial (e.g., a manual choice of color). That large composite network can be rendered reversibly weak or inoperative by cooling either inferotemporal or dorsolateral prefrontal cortex, whereby the retention of the stimulus fails and, with it the performance of the task (Bauer and Fuster, 1976; Fuster et al., 1981). The animal can perform the task but appears incapable of retaining the stimulus of each trial and of carrying out the proper manual choices.

The interaction between the two cortices — inferotemporal and prefrontal — in visual memory can be further substantiated by cooling one cortex and recording with microelectrodes from the other during performance of a delayed matching-to-sample task (Fuster et al., 1985). Prefrontal cooling appears to diminish the capacity of inferotemporal cells to retain colors in short-term memory. Inferotemporal cooling seems to have a similar effect on prefrontal cells. In sum, it appears that by cooling either cortex we interrupt loops of recurrent activation functionally interconnecting the two cortices in the retention of visual short-term memory for a prospective act. The flow of excitatory influences in those loops would be what allows the active retention of the memorandum.

3. Recurrent activation of a memory network

The concept that short-term memory consists in the reverberation of impulses through recurrent networks was first proposed by Hebb (1949). He based that concept on the histological evidence of profuse recurrent fiber connections in the cerebral cortex (Lorente de Nó, 1949), and applied the concept to the role of peristriate cortex in visual short-term memory. The results of our cooling experiments, however, suggest that the recurrent reverberation of visual working memory is more wide-ranging than Hebb envisioned. The sustained activation of a visual memory network seems to require wide-ranging cortico—cortical connectivity as well as the local reentrant circuit in order to test the reverberation idea in active memory, we have used several methods. The first was to construct a spiking computer network model and to train it to retain information for the short term (Zipser et al., 1993). The purpose was to find out if, in short-term memory, the units of the model behaved like real cortical cells in an active memory network. The architecture of the model was essentially reentrant: any unit in the network was connected to all others by reentrant connections (Fig. 1). The model was trained by the backpropagation method (Rumelhart et al., 1986). This method is an error-reducing procedure that, through successive iterations, allows the network to adjust its synaptic weights in order to maintain a stable input-output relationship despite fluctuations in input value. After training, the weights stay fixed.

At variance with conventional backpropagation models, ours has a load signal or gate. When the gate is open (load signal 0), the last input of a given value is allowed in the network and held within it at that value (In the brain, the gate could be a limbic structure or the prefrontal cortex). Allowing for the stochastic firing of real neurons, all input and output values were assumed to reflect spiking probability (Amit, 1990). This assumes that in the absence of change of input or output a cell will discharge at random with fixed probability. Input changes are translated into changes of spike frequency, each with a random distribution inter-spike intervals. In the trained model, single units are substituted by pools of units representing cell assemblies or network components (‘netlets’) in the real brain. The input to a pool of cells is determined by two vectors: one is a function of the total outputs from the network’s pools’, and the other is determined by external inputs. Two additional vectors are the weights of connection with other pools and with external inputs.

On the fully trained network, a memory task trial can be simulated by loading an analog input, that is, the memorandum, and by holding the gate open (load signal 0) throughout the memory period or delay until the act of recognition, when the load signal changes level, in effect closing again the gate without input. In these conditions,
the units of the network behave like cells in the real cortex during the trials of a memory task, especially if a certain amount of internal noise is injected into the network (cells in a real cortical network are also in a noisy environment). The behavior of output cells is unremarkable, since they reflect the transfer function from input to output, defined beforehand. Truly noteworthy, however, is the behavior of the internal units of the network (‘hidden units’) which — with adequate scaling — show temporal patterns of firing that are extraordinarily similar to those of cortical cells in the memory task (Fig. 2). Those patterns of discharge, which so much resemble those of real cells, are part of a repertoire obtained by repeated test of the sample-and-hold operation of the model and generated by the model’s internal architecture.

Consequently, the firing patterns of cortical neurons in a memory task can be understood as a
result of the activation of a fully trained, recurrent memory network with preestablished synaptic weights. The training mechanism by which the weights have been fixed in the model is immaterial to the validity of this conclusion. Of course, backpropagation has little to do with the selective or hebbian mechanisms guiding the storage of memory in the cerebral cortex. What is important is that, once the weights have been set (that is, the memory has been acquired), the short-term activation of the network elicits in its component units similar patterns of firing in the model as in the brain. Thus, the importance of recurrence in the cortical dynamics of working memory is substantiated by the behavior of the hidden units in a model which has recurrence as an essential feature of its functional architecture.

Further, the model we constructed showed another characteristic of exceptional interest for understanding the firing patterns of cortical cells in an active memory network. It showed on close analysis that, in the absence of new inputs or changes in current input, the firing of the network’s units will not remain invariable, but will drift toward one or several frequencies and will shift between them. A unit will change repeatedly and in a more or less regular manner between several frequencies. This phenomenon is most evident at fine temporal resolutions, as changes occur rapidly. Those fine firing shifts are not apparent at coarse temporal scale or in smoothed frequency histograms.

That phenomenon of the model at a fine temporal resolution, despite its unclear origin, obliged us to examine in detail the discharge of real cells of the inferotemporal cortex in a visual memory task. Analyzing spike trains at fine temporal resolution, we discovered rapid changes between alternating frequencies that were similar to those exhibited by the model. We reasoned that those changes might be inherent in recurrent networks and resulted precisely from their reentry property. Cowan (1971) was one of those showing for the first time that recurrent networks of nonlinear units, like ours, tend to drift toward and between certain frequencies called ‘attractors’.

It is unclear whether there are any physiological or anatomical constraints, instead of or in addition to recurrence, that could account for what has been called ‘attractor dynamics’ in cortical networks engaged in the active retention of information. It is possible that attractor dynamics is at least in part determined by intrinsic mechanisms of cortical cells (Llinás, 1990). It seems plausible, however, to explore the physiological role of recurrent brain circuitry as the basis for attractor dynamics in active short-term memory. Supporting this presumption is our evidence of the cortico-cortical loops in memory (Fuster et al., 1985) which I have mentioned above. Also
supporting the hypothesis is the remarkable similarity of firing patterns, which I have also mentioned, between the units of the monkey and those of an artificial recurrent network during active short-term memory.

Assuming that a memory network links the associated attributes of a memorandum, and if these attributes are represented in separate but interconnected cell assemblies of the cortex, it is possible that the activation of the memorandum (i.e. the network) results in the reverberation of impulses between those assemblies. That reverberation could be the foundation of attractor dynamics in active memory. A neuron’s attractor frequency may reflect the circulation of impulses between it and other neurons of the network. Each of the cell’s attractor frequencies might reflect its functional linkage to a different component of the memorandum. Consequently, the cell would be subject to as many attractors as there are associated features of the memorandum. That functional linkage by reentry would be a kind of binding in memory, similar to the binding of objects in perception.

These propositions are testable electrophysiologically. In accord with the principle of attractor dynamics, we tested the general prediction that any given cell in a memory network will show more fluctuations of firing frequency while the network is engaged in active memory than when it is not. That prediction is based on the assumption that, as a member of an active network, the cell will be subject to more inputs — excitatory as well as inhibitory — than when the network is in the inactive state. Since cellular action potentials are generated and inhibited discontinuously as the result of the temporal summation of presynaptic potentials, the more inputs of diverse origin a cell receives the less stable its firing will be. This is especially likely in neurons of a cortical network representing a complex memorandum, and thus subject to many attractors.

For the purpose of testing that prediction, cells of the somatosensory cortex of the monkey were investigated during active memory of the surface features of an object perceived by active touch (Zhou and Fuster, 1996; Bodner et al., 1997). Single-cell firing was recorded as the animal performed a haptic memory task. Each task trial began with the blind palpation of a sample rod with a special surface feature (smooth or rough surface, horizontal or vertical edges). A delay of 12 s followed (memory period), at the end of which the monkey was allowed to palpate two rods and choose, for a reward, the one of the two that was identical to the sample. Two spike trains were selected on every trial from each cell: the first in the 12-s baseline period preceding the trial, and the second in the 12-s memory period.

‘Memory cells’ were found in somatosensory cortex, as in other cortical regions (Fuster, 1995). Some show higher firing frequency as soon as the animal has to retain a stimulus in short-term memory (the delay). Some show a different level of firing depending on the particular stimulus in memory. In our analysis, however, we included cells whose average firing during memorization did not differ significantly from baseline (spontaneous firing between trials). Our aim was to uncover differences in the number of rapid fluctuations of firing frequency (number of attractors?) between baseline and memorization periods.

Although the analysis of inferotemporal cells had provided us with some general idea of the velocity of ‘attractor shifts’ (Zipser et al., 1993), here in parietal cortex we had no notion, a priori, of that velocity and, consequently, of the temporal resolution needed to explore it in a massive body of spike trains from many neurons. Further, we had no grounds to assume that all cells in our sample would show frequency fluctuations, much less at the same resolution. We needed a multiple filter of cell-firing transients.

A ‘filter’ of this kind is, for example, the binary mapping of discrete temporal events — i.e. cell spikes (Fig. 3). This is the method we adopted. It is useful for detecting, in a given time series, fluctuations of frequency (transitions) without prior notion of the range in which they occur. The following mapping procedure was conducted on each of the spike trains from a somatosensory cell during intertrial baseline and ‘delay’ periods. First, the 12-s period was segmented into equal-size bins, and second, a 1 or a 0 was assigned to each bin depending on whether it contained any spike or not. In the resulting binary curve, a frequency
transition was defined as a shift from 0 to 1 or vice versa. By segmenting successively the record into bins of various sizes, the method becomes essentially a non-linear low-pass filter of spike-frequency changes or transitions. We systematically used all bin sizes between 1 and 140 ms in 1-ms increments; in other words, we applied to each spike train 140 filters or time resolutions.

The number of transitions was found to be higher, and at more bin sizes, during delays (memory periods) than during inter-trial baseline periods (Fig. 4). As predicted, the cells appeared subject to increased inputs while the monkey memorized the palpated object. Furthermore, the increase in transitions need not be accompanied by average frequency change. The differences in transitions between baseline and delay were most prominent at bin-sizes between 20 and 50 ms.

As expected, therefore, cortical cells in active memory appear to fluctuate more often between firing frequencies than in spontaneous baseline condition. The transition analysis, however, does not specify precisely those frequencies. Transition differences within a given range of resolution simply imply that the transitions themselves occur most often at the frequencies corresponding to that range, but they do not specify the cell’s discharge between transitions, except within wide limits — from 10 to 50 Hz. In a given cell, an increase of transitions between attractor frequencies would indicate the multiplication of reentries upon the cell from within the memory network to which the cell belongs. The attractors, however, need not be ‘fixed-point’ (fixed frequencies). Complex patterns of transitions (‘limit-cycle attractors’) are also possible. These would be even more consistent with multiple reentry. We are now investigating such patterns.

In conclusion, the excitatory reentry through recurrent circuits appears to be a plausible mechanism for maintaining the activation of a memory network. That reentry may functionally bind
Fig. 4. Frequency and transition analysis of spike trains (all 12-s long) from two parietal cells in all trials of a haptic memory task during recording (cell positions marked by a small triangle in cortical cuts below). Upper graphs: Ranking of spike trains by average frequency (trains from intertrial baseline in white, from delay in black; trains are ranked from left to right in descending order, independent of trial sequence or period-baseline or delay). Lower graphs: Ranking of the same trains by transition counts on binary curves with a mapping bin size of 20 ms for cell A and 23 ms for cell B (the mapping bins are selected for illustration of clear ranking separation between baseline and delay). Note that in both cells transition counts were higher for delay than baseline ($P < 0.01$) in the absence of significant differences in average spike frequency. From Bodner et al., 1997.

together the associated components of the memory represented by the network. The reentrant binding may take place within a relatively discrete cortical domain or between widely separated cortical areas, depending on the dispersion of the associated features of the active memory.

References


