Jianfeng Feng

Computational Neuroscience: A Comprehensive Approach

CRC PRESS
Boca Raton Ann Arbor London Tokyo

Chapter 16

The Operation of Memory Systems in the Brain

Edmund T. Rolls
University of Oxford, Dept. of Experimental Psychology, South Parks Road, Oxford OX1 3UD, England. www.cns.ox.ac.uk

0-8493-0052-5/00/$0.00+$.50

16.1 Introduction

This chapter describes memory systems in the brain based on closely linked neurobiological and computational approaches. The neurobiological approaches include evidence from brain lesions which show the type of memory for which each of the brain systems considered is necessary; and analysis of neuronal activity in each of these systems to show what information is represented in them, and the changes that take place during learning. Much of the neurobiology considered is from non-human primates as well as humans, because the operation of some of the brain systems involved in memory and connected to them have undergone great development in primates. Some such brain systems include those in the temporal lobe, which develops massively in primates for vision, and which sends inputs to the hippocampus via highly developed parahippocampal regions; and the prefrontal cortex. Many memory systems in primates receive outputs from the primate inferior temporal visual cortex, and understanding the perceptual representations in this of objects, and how they are appropriate as inputs to different memory systems, helps to provide a coherent way to understand the different memory systems in the brain (see [82], which provides a more extensive treatment of the brain architectures used for perception and memory). The computational approaches are essential in order to understand how the circuitry could retrieve as well as store memories, the capacity of each memory system in the brain, the interactions between memory and perceptual systems, and the speed of operation of the memory systems in the brain.

The architecture, principles of operation, and properties of the main types of network referred to here, autoassociation or attractor networks, pattern association networks, and competitive networks, are described by [82] and [92].

16.2 Functions of the hippocampus in long-term memory

The inferior temporal visual cortex projects via the perirhinal cortex and entorhinal cortex to the hippocampus (see Figure 16.1), which is implicated in long term memory, of, for example, where objects are located in spatial scenes, which can be thought of as an example of episodic memory. The architecture shown in Figure 16.1 indicates that the hippocampus provides a region where visual outputs from the inferior temporal visual cortex can, via the perirhinal cortex and entorhinal cortex, be brought together with outputs from the ends of other cortical processing streams. In this section, we consider how the visual input about objects is in the correct form for the types of memory implemented by the perirhinal and hippocampal systems, how the hippocampus of primates contains a representation of the visual space being viewed, how this may be similar computationally to the apparently very different

representation of places that is present in the rat hippocampus, how these spatial representations are in a form that could be implemented by a continuous attractor which could be updated in the dark by idiothetic inputs, and how a unified attractor theory of hippocampal function can be formulated using the concept of mixed attractors. The visual output from the inferior temporal visual cortex may be used to provide the perirhinal and hippocampal systems with information about objects that is useful in visual recognition memory, in episodic memory of where objects are seen, and for building spatial representations of visual scenes. Before summarizing the computational approaches to these issues, we first summarize some of the empirical evidence that needs to be accounted for in computational models.

16.2.1 Effects of damage to the hippocampus and connected structures on object-place and episodic memory

Partly because of the evidence that in humans with bilateral damage to the hippocampus and nearby parts of the temporal lobe, anterograde amnesia is produced [100], there is continuing great interest in how the hippocampus and connected structures operate in memory. The effects of damage to the hippocampus indicate that the very long-term storage of at least some types of information is not in the hippocampus, at least in humans. On the other hand, the hippocampus does appear to be necessary to learn certain types of information, that have been characterized as declarative, or knowing that, as contrasted with procedural, or knowing how, which is spared in amnesia. Declarative memory includes what can be declared or brought to mind as a proposition or an image. Declarative memory includes episodic memory (memory for particular episodes), and semantic memory (memory for facts) [100].

In monkeys, damage to the hippocampus or to some of its connections such as the fornix produces deficits in learning about where objects are and where responses must be made (see [12]) and [76]. For example, macaques and humans with damage to the hippocampus or fornix are impaired in object-place memory tasks in which not only the objects seen, but where they were seen, must be remembered [28, 60, 99]. Such object-place tasks require a whole-scene or snapshot-like memory [25]. Also, fornix lesions impair conditional left-right discrimination learning, in which the visual appearance of an object specifies whether a response is to be made to the left or the right [94]. A comparable deficit is found in humans [61]. Fornix sectioned monkeys are also impaired in learning on the basis of a spatial cue which object to choose (e.g. if two objects are on the left, choose object A, but if the two objects are on the right, choose object B) [26]. Further, monkeys with fornix damage are also impaired in using information about their place in an environment. For example, [27] found learning impairments when the position of the monkey in the room determined which of two or more objects the monkey had to choose. Rats with hippocampal lesions are impaired in using environmental spatial cues to remember particular places [35, 45], and it has been argued that the necessity to utilize allocentric spatial cues [14], to utilize spatial cues or bridge delays [34, 37], or to perform relational operations on

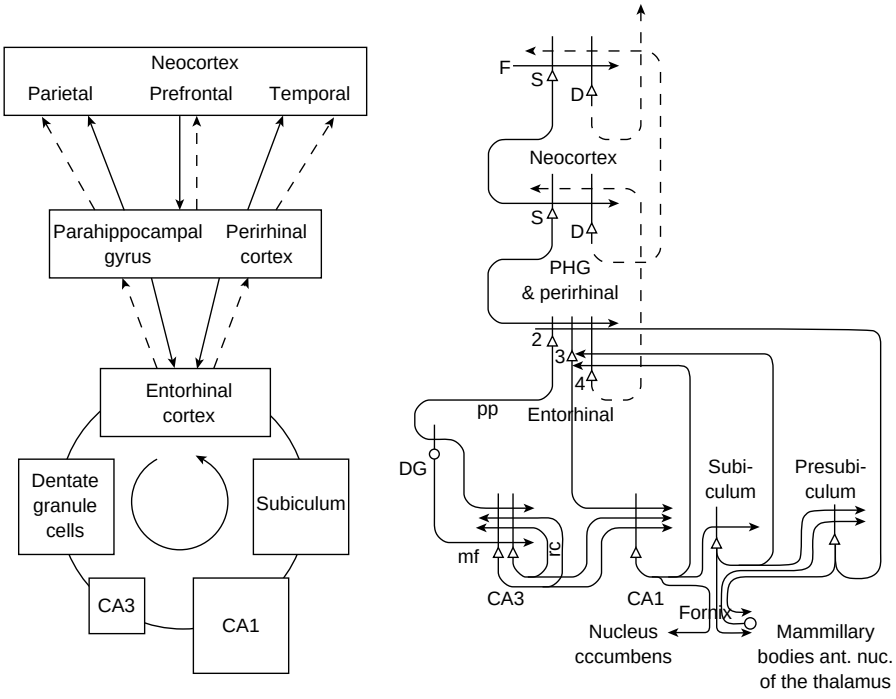

Figure 16.1

Forward connections (solid lines) from areas of cerebral association neocortex via the parahippocampal gyrus and perirhinal cortex, and entorhinal cortex, to the hippocampus; and backprojections (dashed lines) via the hippocampal CA1 pyramidal cells, subiculum, and parahippocampal gyrus to the neocortex. There is great convergence in the forward connections down to the single network implemented in the CA3 pyramidal cells; and great divergence again in the backprojections. Left: block diagram. Right: more detailed representation of some of the principal excitatory neurons in the pathways. Abbreviations: D, Deep pyramidal cells; DG, dentate granule cells; F, forward inputs to areas of the association cortex from preceding cortical areas in the hierarchy. mf: mossy fibres; PHG, parahippocampal gyrus and perirhinal cortex; pp, perforant path; rc, recurrent collaterals of the CA3 hippocampal pyramidal cells; S, superficial pyramidal cells; 2, pyramidal cells in layer 2 of the entorhinal cortex; 3, pyramidal cells in layer 3 of the entorhinal cortex; 5, 6, pyramidal cells in the deep layers of the entorhinal cortex. The thick lines above the cell bodies represent the dendrites.

remembered material [19], may be characteristic of the deficits.

One way of relating the impairment of spatial processing to other aspects of hippocampal function (including the memory of recent events or episodes in humans) is to note that this spatial processing involves a snapshot type of memory, in which one whole scene with its often unique set of parts or elements must be remembered. This memory may then be a special case of episodic memory, which involves an arbitrary association of a set of spatial and/or non-spatial events that describe a past episode. For example, the deficit in paired associate learning in humans (see [100]) may be especially evident when this involves arbitrary associations between words, for example, window — lake.

It appears that the deficits in 'recognition' memory (tested for example for visual stimuli seen recently in a delayed match to sample task) produced by damage to this brain region are related to damage to the perirhinal cortex [122, 123], which receives from high order association cortex and has connections to the hippocampus (see Figure 16.1) [107, 108]. The functions of the perirhinal cortex in memory are discussed by [82].

16.2.2 Neurophysiology of the hippocampus and connected areas

In the rat, many hippocampal pyramidal cells fire when the rat is in a particular place, as defined for example by the visual spatial cues in an environment such as a room [39, 53, 54]. There is information from the responses of many such cells about the place where the rat is in the environment. When a rat enters a new environment B connected to a known environment A, there is a period in the order of 10 minutes in which as the new environment is learned, some of the cells that formerly had place fields in A develop instead place fields in B. It is as if the hippocampus sets up a new spatial representation which can map both A and B, keeping the proportion of cells active at any one time approximately constant [118]. Some rat hippocampal neurons are found to be more task-related, responding for example to olfactory stimuli to which particular behavioural responses must be made [19], and some of these neurons may in different experiments show place-related responses.

It was recently discovered that in the primate hippocampus, many spatial cells have responses not related to the place where the monkey is, but instead related to the place where the monkey is looking [78, 79, 85]. These are called 'spatial view cells', an example of which is shown in Figure 16.2. These cells encode information in allocentric (world-based, as contrasted with egocentric, body-related) coordinates [29, 93]. They can in some cases respond to remembered spatial views in that they respond when the view details are obscured, and use idiothetic (self-motion) cues including eye position and head direction to trigger this memory recall operation [71]. Another idiothetic input that drives some primate hippocampal neurons is linear and axial whole body motion [58], and in addition, the primate presubiculum has been shown to contain head direction cells [72].

Part of the interest of spatial view cells is that they could provide the spatial repre-

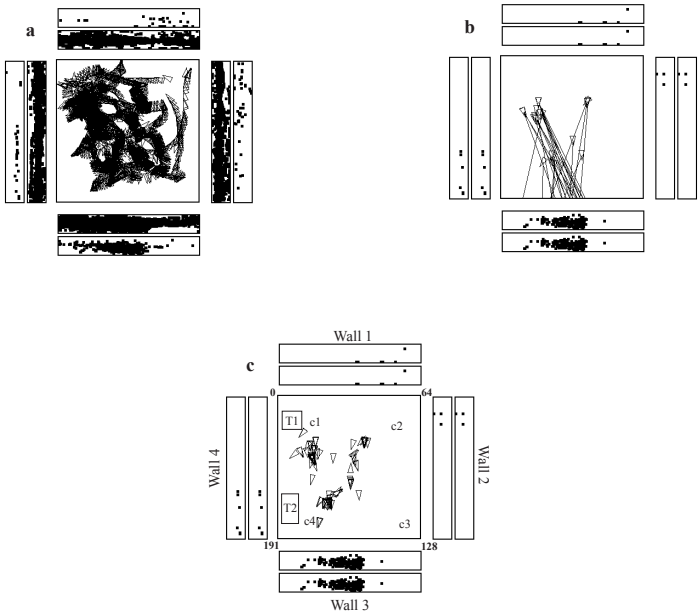

Figure 16.2

Examples of the firing of a hippocampal spatial view cell when the monkey was walking around the laboratory. a. The firing of the cell is indicated by the spots in the outer set of 4 rectangles, each of which represents one of the walls of the room. There is one spot on the outer rectangle for each action potential. The base of the wall is towards the centre of each rectangle. The positions on the walls fixated during the recording sessions are indicated by points in the inner set of 4 rectangles, each of which also represents a wall of the room. The central square is a plan view of the room, with a triangle printed every 250 ms to indicate the position of the monkey, thus showing that many different places were visited during the recording sessions. b. A similar representation of the same 3 recording sessions as in (a), but modified to indicate some of the range of monkey positions and horizontal gaze directions when the cell fired at more than 12 spikes/s. c. A similar representation of the same 3 recording sessions as in (b), but modified to indicate more fully the range of places when the cell fired. The triangle indicates the current position of the monkey, and the line projected from it shows which part of the wall is being viewed at any one time while the monkey is walking. One spot is shown for each action potential. (After Georges-François, Rolls and Robertson, 1999)

sentation required to enable primates to perform object-place memory, for example remembering where they saw a person or object, which is an example of an episodic memory, and indeed similar neurons in the hippocampus respond in object-place

memory tasks [84]. Associating together such a spatial representation with a representation of a person or object could be implemented by an autoassociation network implemented by the recurrent collateral connections of the CA3 hippocampal pyramidal cells [75, 76, 92]. Some other primate hippocampal neurons respond in the object-place memory task to a combination of spatial information and information about the object seen [84]. Further evidence for this convergence of spatial and object information in the hippocampus is that in another memory task for which the hippocampus is needed, learning where to make spatial responses conditional on which picture is shown, some primate hippocampal neurons respond to a combination of which picture is shown, and where the response must be made [13, 48].

These primate spatial view cells are thus unlike place cells found in the rat [39, 51, 53, 54, 118]. Primates, with their highly developed visual and eye movement control systems, can explore and remember information about what is present at places in the environment without having to visit those places. Such spatial view cells in primates would thus be useful as part of a memory system, in that they would provide a representation of a part of space that would not depend on exactly where the monkey or human was, and that could be associated with items that might be present in those spatial locations. An example of the utility of such a representation in humans would be remembering where a particular person had been seen. The primate spatial representations would also be useful in remembering trajectories through environments, of use for example in short-range spatial navigation [58, 79].

The representation of space in the rat hippocampus, which is of the place where the rat is, may be related to the fact that with a much less developed visual system than the primate, the rat's representation of space may be defined more by the olfactory and tactile as well as distant visual cues present, and may thus tend to reflect the place where the rat is. An interesting hypothesis on how this difference could arise from essentially the same computational process in rats and monkeys is as follows [17, 79]. The starting assumption is that in both the rat and the primate, the dentate granule cells and the CA3 and CA1 pyramidal cells respond to combinations of the inputs received. In the case of the primate, a combination of visual features in the environment will over a typical viewing angle of perhaps 10–20 degrees result in the formation of a spatial view cell, the effective trigger for which will thus be a combination of visual features within a relatively small part of space. In contrast, in the rat, given the very extensive visual field which may extend over 180–270 degrees, a combination of visual features formed over such a wide visual angle would effectively define a position in space, that is a place. The actual processes by which the hippocampal formation cells would come to respond to feature combinations could be similar in rats and monkeys, involving for example competitive learning in the dentate granule cells, autoassociation learning in CA3 pyramidal cells, and competitive learning in CA1 pyramidal cells [75, 76, 92, 116]. Thus spatial view cells in primates and place cells in rats might arise by the same computational process but be different by virtue of the fact that primates are foveate and view a small part of the visual field at any one time, whereas the rat has a very wide visual field. Although the representation of space in rats therefore may be in some ways analogous to the representation of space in the primate hippocampus, the difference does have

implications for theories, and modelling, of hippocampal function.
In rats, the presence of place cells has led to theories that the rat hippocampus is a spatial cognitive map, and can perform spatial computations to implement navigation through spatial environments [11, 10, 54, 57]. The details of such navigational theories could not apply in any direct way to what is found in the primate hippocampus. Instead, what is applicable to both the primate and rat hippocampal recordings is that hippocampal neurons contain a representation of space (for the rat, primarily where the rat is, and for the primate primarily of positions 'out there' in space) which is a suitable representation for an episodic memory system. In primates, this would enable one to remember, for example, where an object was seen. In rats, it might enable memories to be formed of where particular objects (for example those defined by olfactory, tactile, and taste inputs) were found. Thus at least in primates, and possibly also in rats, the neuronal representation of space in the hippocampus may be appropriate for forming memories of events (which usually in these animals have a spatial component). Such memories would be useful for spatial navigation, for which according to the present hypothesis the hippocampus would implement the memory component but not the spatial computation component. Evidence that what neuronal recordings have shown is represented in the non-human primate hippocampal system may also be present in humans is that regions of the hippocampal formation can be activated when humans look at spatial views [21, 55].

16.2.3 Hippocampal models

These neuropsychological and neurophysiological analyses are complemented by neuronal network models of how the hippocampus could operate to store and retrieve large numbers of memories [73, 75, 76, 92, 115, 116]). One key hypothesis (adopted also by [46]) is that the hippocampal CA3 recurrent collateral connections which spread throughout the CA3 region provide a *single autoassociation network* that enables the firing of *any* set of CA3 neurons representing one part of a memory to be associated together with the firing of any other set of CA3 neurons representing another part of the same memory (cf. [44]). The generic architecture of an attractor network is shown in Figure 16.5. Associatively modifiable synapses in the recurrent collateral synapses allow memories to be stored, and then later retrieved from only a part, as described by [4, 33, 32, 82, 92]. The number of patterns p each representing a different memory that could be stored in the CA3 system operating as an autoassociation network would be as shown in equation 16.1 (see [82, 92], which describe extensions to the analysis developed by [33]).

$$p \approx \frac{C^{\mathrm{RC}}}{a \ln(\frac{1}{a})} k \tag{16.1}$$

where C^{RC} is the number of synapses on the dendrites of each neuron devoted to the recurrent collaterals from other CA3 neurons in the network, a is the sparseness of the representation, and k is a factor that depends weakly on the detailed structure

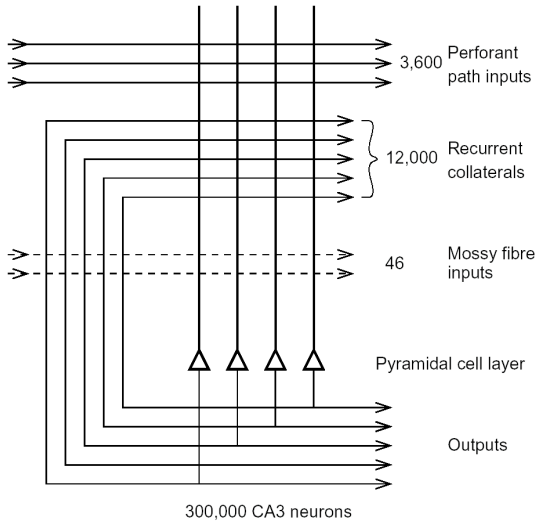

Figure 16.3

The numbers of connections from three different sources onto each CA3 cell from three different sources in the rat. (After Treves and Rolls 1992, and Rolls and Treves 1998.)

of the rate distribution, on the connectivity pattern, etc., but is roughly in the order of 0.2–0.3. Given that C^{RC} is approximately 12,000 in the rat, the resulting storage capacity would be greater than 12,000 memories, and perhaps up to 36,000 memories if the sparseness a of the representation was as low as 0.02 [115, 116].

Another part of the hypothesis is that the very sparse (see Figure 16.3) but powerful connectivity of the mossy fibre inputs to the CA3 cells from the dentate granule cells is important during learning (but not recall) to force a new, arbitrary, set of firing onto the CA3 cells which dominates the activity of the recurrent collaterals, so enabling a new memory represented by the firing of the CA3 cells to be stored [73, 75, 115].

The perforant path input to the CA3 cells, which is numerically much larger but at the apical end of the dendrites, would be used to initiate recall from an incomplete pattern [92, 115]. The prediction of the theory about the necessity of the mossy fibre inputs to the CA3 cells during learning but not recall has now been confirmed [42]. A way to enhance the efficacy of the mossy fibre system relative to the CA3 recurrent collateral connections during learning may be to increase the level of acetyl choline by increasing the firing of the septal cholinergic cells [31].

Another key part of the quantitative theory is that not only can retrieval of a memory by an incomplete cue be performed by the operation of the associatively modified CA3 recurrent collateral connections, but also that recall of that information to the neocortex can be performed via CA1 and the hippocampo-cortical and cortico-cortical backprojections [76, 81, 92, 116] shown in Figure 16.1. In this case, the

number of memory patterns p^{BP} that can be retrieved by the backprojection system is

$$p^{\text{BP}} \approx \frac{C^{\text{BP}}}{a^{\text{BP}} \ln(\frac{1}{a^{\text{BP}}})} k^{\text{BP}} \tag{16.2}$$

where C^{BP} is the number of synapses on the dendrites of each neuron devoted to backprojections from the preceding stage (dashed lines in Figure 16.1), a^{BP} is the sparseness of the representation in the backprojection pathways, and k^{BP} is a factor that depends weakly on the detailed structure of the rate distribution, on the connectivity pattern, etc., but is roughly in the order of 0.2–0.3. The insight into this quantitative analysis came from treating each layer of the backprojection hierarchy as being quantitatively equivalent to another iteration in a single recurrent attractor network [114, 116]. The need for this number of connections to implement recall, and more generally constraint satisfaction in connected networks (see [82]), provides a fundamental and quantitative reason for why there are approximately as many backprojections as forward connections between the adjacent connected cortical areas in a cortical hierarchy. This, and other computational approaches to hippocampal function, are included in a special issue of the journal Hippocampus (1996), 6(6).

Another aspect of the theory is that the operation of the CA3 system to implement recall, and of the backprojections to retrieve the information, would be sufficiently fast, given the fast recall in associative networks built of neurons with continuous dynamics (see [82]).

16.2.4 Continuous spatial representations, path integration, and the use of idiothetic inputs

The fact that spatial patterns, which imply continuous representations of space, are represented in the hippocampus has led to the application of continuous attractor models to help understand hippocampal function. Such models have been developed by [8, 95, 101, 102, 104, 105], (see [82]). Indeed, we have shown how a continuous attractor network could enable the head direction cell firing of presubicular cells to be maintained in the dark, and updated by idiothetic (self-motion) head rotation cell inputs [72, 101]. The continuous attractor model has been developed to understand how place cell firing in rats can be maintained and updated by idiothetic inputs in the dark [104]. The continuous attractor model has also been developed to understand how spatial view cell firing in primates can be maintained and updated by idiothetic eye movement and head direction inputs in the dark [71, 105].

The way in which path integration could be implemented in the hippocampus or related systems is described next. Single-cell recording studies have shown that some neurons represent the current position along a continuous physical dimension or space even when no inputs are available, for example in darkness. Examples include neurons that represent the positions of the eyes (i.e., eye direction with respect to the head), the place where the animal is looking in space, head direction, and the

place where the animal is located. In particular, examples of such classes of cells include head direction cells in rats [50, 62, 110, 111] and primates [72], which respond maximally when the animal's head is facing in a particular preferred direction; place cells in rats [43, 47, 49, 52, 56] that fire maximally when the animal is in a particular location; and spatial view cells in primates that respond when the monkey is looking towards a particular location in space [29, 71, 85]. In the parietal cortex there are many spatial representations, in several different coordinate frames (see [6] and [82]), and they have some capability to remain active during memory periods when the stimulus is no longer present. Even more than this, the dorsolateral prefrontal cortex networks to which the parietal networks project have the capability to maintain spatial representations active for many seconds or minutes during short term memory tasks, when the stimulus is no longer present (see below).

A class of network that can maintain the firing of its neurons to represent any location along a continuous physical dimension such as spatial position, head direction, etc is a 'Continuous Attractor' neural network (CANN). It uses excitatory recurrent collateral connections between the neurons to reflect the distance between the neurons in the state space of the animal (e.g. head direction space). These networks can maintain the bubble of neural activity constant for long periods wherever it is started to represent the current state (head direction, position, etc) of the animal, and are likely to be involved in many aspects of spatial processing and memory, including spatial vision. Global inhibition is used to keep the number of neurons in a bubble or packet of actively firing neurons relatively constant, and to help to ensure that there is only one activity packet. Continuous attractor networks can be thought of as very similar to autoassociation or discrete attractor networks (see [82]), and have the same architecture, as illustrated in Figure 16.5. The main difference is that the patterns stored in a CANN are continuous patterns, with each neuron having broadly tuned firing which decreases with for example a Gaussian function as the distance from the optimal firing location of the cell is varied, and with different neurons having tuning that overlaps throughout the space. Such tuning is illustrated in Figure 16.4. For comparison, autoassociation networks normally have discrete (separate) patterns (each pattern implemented by the firing of a particular subset of the neurons), with no continuous distribution of the patterns throughout the space (see Figure 16.4). A consequent difference is that the CANN can maintain its firing at any location in the trained continuous space, whereas a discrete attractor or autoassociation network moves its population of active neurons towards one of the previously learned attractor states, and thus implements the recall of a particular previously learned pattern from an incomplete or noisy (distorted) version of one of the previously learned patterns. The energy landscape of a discrete attractor network (see [82]) has separate energy minima, each one of which corresponds to a learned pattern, whereas the energy landscape of a continuous attractor network is flat, so that the activity packet remains stable with continuous firing wherever it is started in the state space. (The state space refers to set of possible spatial states of the animal in its environment, e.g. the set of possible head directions.) I next describe the operation and properties of continuous attractor networks, which have been studied by for example [3], [112], and [120], and then, following [101], address four key issues about the biological

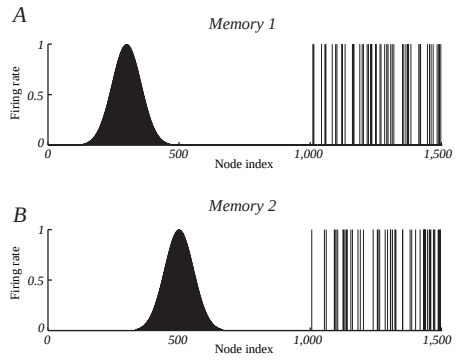

Figure 16.4

The types of firing patterns stored in continuous attractor networks are illustrated for the patterns present on neurons 1–1000 for Memory 1 (when the firing is that produced when the spatial state represented is that for location 300), and for Memory 2 (when the firing is that produced when the spatial state represented is that for location 500). The continuous nature of the spatial representation results from the fact that each neuron has a Gaussian firing rate that peaks at its optimal location. This particular mixed network also contains discrete representations that consist of discrete subsets of active binary firing rate neurons in the range 1001–1500. The firing of these latter neurons can be thought of as representing the discrete events that occur at the location. Continuous attractor networks by definition contain only continuous representations, but this particular network can store mixed continuous and discrete representations, and is illustrated to show the difference of the firing patterns normally stored in separate continuous attractor and discrete attractor networks. For this particular mixed network, during learning, Memory 1 is stored in the synaptic weights, then Memory 2, etc, and each memory contains part that is continuously distributed to represent physical space, and part that represents a discrete event or object.

application of continuous attractor network models.

One key issue in such continuous attractor neural networks is how the synaptic strengths between the neurons in the continuous attractor network could be learned in biological systems (Section 16.2.4.2).

A second key issue in such Continuous Attractor neural networks is how the bubble of neuronal firing representing one location in the continuous state space should be updated based on non-visual cues to represent a new location in state space. This is essentially the problem of path integration: how a system that represents a memory of where the agent is in physical space could be updated based on idiothetic (self-motion) cues such as vestibular cues (which might represent a head velocity signal), or proprioceptive cues (which might update a representation of place based on movements being made in the space, during for example walking in the dark).

A third key issue is how stability in the bubble of activity representing the current location can be maintained without much drift in darkness, when it is operating as a memory system (see [82] and [101]).
A fourth key issue is considered below in which I describe networks that store both continuous patterns and discrete patterns (see Figure 16.4), which can be used to store for example the location in (continuous, physical) space where an object (a discrete item) is present.

16.2.4.1 The generic model of a continuous attractor network

The generic model of a continuous attractor is as follows. (The model is described in the context of head direction cells, which represent the head direction of rats [50, 110] and macaques [72], and can be reset by visual inputs after gradual drift in darkness.) The model is a recurrent attractor network with global inhibition. It is different from a Hopfield attractor network [33] primarily in that there are no discrete attractors formed by associative learning of discrete patterns. Instead there is a set of neurons that are connected to each other by synaptic weights w_{ij} that are a simple function, for example Gaussian, of the distance between the states of the agent in the physical world (e.g., head directions) represented by the neurons. Neurons that represent similar states (locations in the state space) of the agent in the physical world have strong synaptic connections, which can be set up by an associative learning rule, as described in Section 16.2.4.2. The network updates its firing rates by the following 'leaky-integrator' dynamical equations. The continuously changing activation h_i^{HD} of each head direction cell i is governed by the Equation

$$\frac{dh_i^{\mathrm{HD}}(t)}{dt} = -h_i^{\mathrm{HD}}(t) + \frac{\phi_0}{C^{\mathrm{HD}}} \sum_j (w_{ij} - w^{\mathrm{inh}}) r_j^{\mathrm{HD}}(t) + I_i^V, \tag{16.3}$$

where r_j^{HD} is the firing rate of head direction cell j, w_{ij} is the excitatory (positive) synaptic weight from head direction cell j to cell i, w^{inh} is a global constant describing the effect of inhibitory interneurons, and τ is the time constant of the system[1]. The term $-h_i^{\mathrm{HD}}(t)$ indicates the amount by which the activation decays (in the leaky integrator neuron) at time t. (The network is updated in a typical simulation at much smaller timesteps than the time constant of the system, τ.) The next term in Equation (16.3) is the input from other neurons in the network r_j^{HD} weighted by the recurrent collateral synaptic connections w_{ij} (scaled by a constant ϕ_0 and C^{HD} which is the number of synaptic connections received by each head direction cell from other head direction cells in the continuous attractor). The term I_i^V represents a visual input to head direction cell i. Each term I_i^V is set to have a Gaussian response profile in most continuous attractor networks, and this sets the firing of the cells in the continuous

[1]Note that here I use r rather than y to refer to the firing rates of the neurons in the network, remembering that, because this is a recurrently connected network (see Figure 16.5), the output from a neuron y_i might be the input x_j to another neuron.

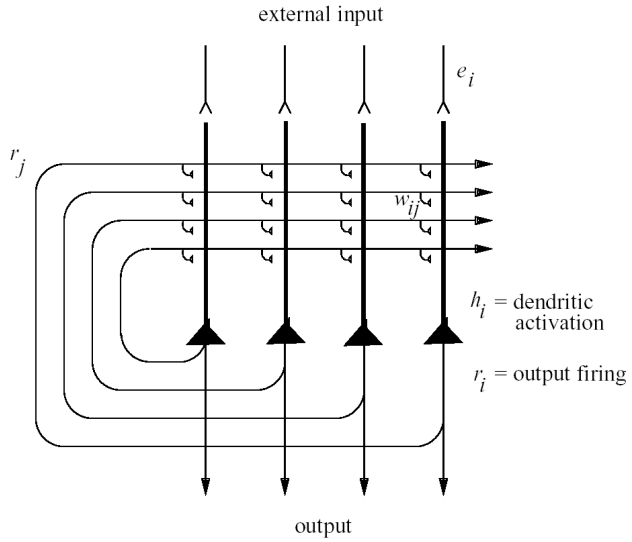

Figure 16.5
The architecture of an attractor neural network.

attractor to have Gaussian response profiles as a function of where the agent is located in the state space (see e.g., Figure 16.4), but the Gaussian assumption is not crucial. (It is known that the firing rates of head direction cells in both rats [50, 110] and macaques [72] is approximately Gaussian.) When the agent is operating without visual input, in memory mode, then the term I_i^V is set to zero. The firing rate r_i^{HD} of cell i is determined from the activation h_i^{HD} and the sigmoid function

$$r_i^{\mathrm{HD}}(t) = \frac{1}{1 + e^{-2\beta(h_i^{\mathrm{HD}}(t) - \alpha)}}, \tag{16.4}$$

where α and β are the sigmoid threshold and slope, respectively.

16.2.4.2 Learning the synaptic strengths between the neurons that implement a continuous attractor network

So far we have said that the neurons in the continuous attractor network are connected to each other by synaptic weights w_{ij} that are a simple function, for example Gaussian, of the distance between the states of the agent in the physical world (e.g. head directions, spatial views etc) represented by the neurons. In many simulations, the weights are set by formula to have weights with these appropriate Gaussian values. However, [101] showed how the appropriate weights could be set up by learning. They started with the fact that since the neurons have broad tuning that may be Gaussian in shape, nearby neurons in the state space will have overlapping spatial fields, and will thus be co-active to a degree that depends on the distance between them.

They postulated that therefore the synaptic weights could be set up by associative learning based on the co-activity of the neurons produced by external stimuli as the animal moved in the state space. For example, head direction cells are forced to fire during learning by visual cues in the environment that produce Gaussian firing as a function of head direction from an optimal head direction for each cell. The learning rule is simply that the weights w_{ij} from head direction cell j with firing rate r_j^{HD} to head direction cell i with firing rate r_i^{HD} are updated according to an associative (Hebb) rule

$$\delta w_{ij} = k r_i^{\mathrm{HD}} r_j^{\mathrm{HD}} \tag{16.5}$$

where δw_{ij} is the change of synaptic weight and k is the learning rate constant. During the learning phase, the firing rate r_i^{HD} of each head direction cell i might be the following Gaussian function of the displacement of the head from the optimal firing direction of the cell

$$r_i^{\mathrm{HD}} = e^{-s_{\mathrm{HD}}^2/2\sigma_{\mathrm{HD}}^2}, \tag{16.6}$$

where s_{HD} is the difference between the actual head direction x (in degrees) of the agent and the optimal head direction x_i for head direction cell i, and σ_{HD} is the standard deviation.

[101] showed that after training at all head directions, the synaptic connections develop strengths that are an almost Gaussian function of the distance between the cells in head direction space, as shown in Figure 16.6 (left). Interestingly if a non-linearity is introduced into the learning rule that mimics the properties of NMDA receptors by allowing the synapses to modify only after strong postsynaptic firing is present, then the synaptic strengths are still close to a Gaussian function of the distance between the connected cells in head direction space (see Figure 16.6, left). They showed that after training, the continuous attractor network can support stable activity packets in the absence of visual inputs (see Figure 16.6, right) provided that global inhibition is used to prevent all the neurons becoming activated. (The exact stability conditions for such networks have been analyzed by [3]). Thus [101] demonstrated biologically plausible mechanisms for training the synaptic weights in a continuous attractor using a biologically plausible local learning rule.

So far, we have considered how spatial representations could be stored in continuous attractor networks, and how the activity can be maintained at any location in the state space in a form of short term memory when the external (e.g. visual) input is removed. However, many networks with spatial representations in the brain can be updated by internal, self-motion (i.e. idiothetic), cues even when there is no external (e.g. visual) input. Examples are head direction cells in the presubiculum of rats and macaques, place cells in the rat hippocampus, and spatial view cells in the primate hippocampus (see Section 16.2). The major question arises about how such idiothetic inputs could drive the activity packet in a continuous attractor network, and in particular, how such a system could be set up biologically by self-organizing learning.

One approach to simulating the movement of an activity packet produced by idiothetic cues (which is a form of path integration whereby the current location is calculated from recent movements) is to employ a look-up table that stores (taking

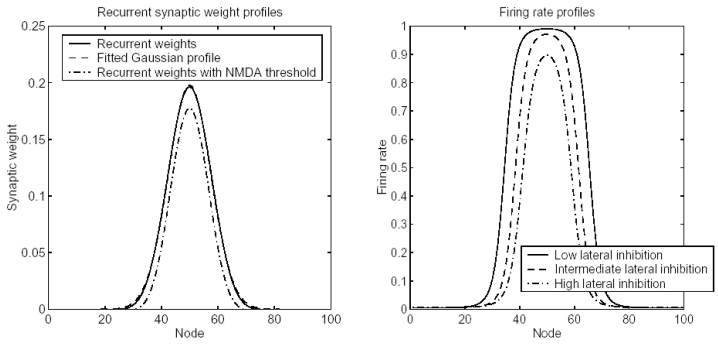

Figure 16.6

Training the weights in a continuous attractor network with an associative rule (equation 16.5). Left: The trained recurrent synaptic weights from head direction cell 50 to the other head direction cells in the network arranged in head direction space (solid curve). The dashed line shows a Gaussian curve fitted to the weights shown in the solid curve. The dash-dot curve shows the recurrent synaptic weights trained with rule equation (16.5), but with a non-linearity introduced that mimics the properties of NMDA receptors by allowing the synapses to modify only after strong postsynaptic firing is present. Right: The stable firing rate profiles forming an activity packet in the continuous attractor network during the testing phase when the training (visual) inputs are no longer present. The firing rates are shown after the network has been initially stimulated by visual input to initialize an activity packet, and then allowed to settle to a stable activity profile without visual input. The three graphs show the firing rates for low, intermediate and high values of the lateral inhibition parameter w^{inh}. For both left and right plots, the 100 head direction cells are arranged according to where they fire maximally in the head direction space of the agent when visual cues are available. After Stringer, Trappenberg, Rolls and de Araujo (2002).

head direction cells as an example), for every possible head direction and head rotational velocity input generated by the vestibular system, the corresponding new head direction [95]. Another approach involves modulating the strengths of the recurrent synaptic weights in the continuous attractor on one but not the other side of a currently represented position, so that the stable position of the packet of activity, which requires symmetric connections in different directions from each node, is lost, and the packet moves in the direction of the temporarily increased weights, although no possible biological implementation was proposed of how the appropriate dynamic synaptic weight changes might be achieved [120]. Another mechanism (for head direction cells) [97] relies on a set of cells, termed (head) rotation cells, which are co-activated by head direction cells and vestibular cells and drive the activity of the attractor network by anatomically distinct connections for clockwise and counterclockwise rotation cells, in what is effectively a look-up table. However, no proposal was made about how this could be achieved by a biologically plausible learning pro-

cess, and this has been the case until recently for most approaches to path integration in continuous attractor networks, which rely heavily on rather artificial pre-set synaptic connectivities.

[101] introduced a proposal with more biological plausibility about how the synaptic connections from idiothetic inputs to a continuous attractor network can be learned by a self-organizing learning process. The essence of the hypothesis is described with Figure 16.7. The continuous attractor synaptic weights w^{RC} are set up under the influence of the external visual inputs I^{V} as described in Section 16.2.4.2. At the same time, the idiothetic synaptic weights w^{ID} (in which the ID refers to the fact that they are in this case produced by idiothetic inputs, produced by cells that fire to represent the velocity of clockwise and anticlockwise head rotation), are set up by associating the change of head direction cell firing that has just occurred (detected by a trace memory mechanism described below) with the current firing of the head rotation cells r^{ID}. For example, when the trace memory mechanism incorporated into the idiothetic synapses w^{ID} detects that the head direction cell firing is at a given location (indicated by the firing r^{HD}) and is moving clockwise (produced by the altering visual inputs I^{V}), and there is simultaneous clockwise head rotation cell firing, the synapses w^{ID} learn the association, so that when that rotation cell firing occurs later without visual input, it takes the current head direction firing in the continuous attractor into account, and moves the location of the head direction attractor in the appropriate direction.

For the learning to operate, the idiothetic synapses onto head direction cell i with firing r_i^{HD} need two inputs: the memory traced term from other head direction cells $\overline{r}_j^{\mathrm{HD}}$ (given by

$$\overline{r}^{\mathrm{HD}}(t+\delta t) = (1-\eta) r^{\mathrm{HD}}(t+\delta t) + \eta \overline{r}^{\mathrm{HD}}(t) \tag{16.7}$$

where η is a parameter set in the interval [0,1] which determines the contribution of the current firing and the previous trace), and the head rotation cell input with firing r_k^{ID}; and the learning rule can be written

$$\delta w_{ijk}^{\mathrm{ID}} = \tilde{k}\, r_i^{\mathrm{HD}}\, \overline{r}_j^{\mathrm{HD}}\, r_k^{\mathrm{ID}}, \tag{16.8}$$

where $\tilde{k}$ is the learning rate associated with this type of synaptic connection. The head rotation cell firing (r_k^{ID}) could be as simple as one set of cells that fire for clockwise head rotation (for which k might be 1), and a second set of cells that fire for anticlockwise head rotation (for which k might be 2).

After learning, the firing of the head direction cells would be updated in the dark (when $I_i^V = 0$) by idiothetic head rotation cell firing r_k^{ID} as follows

$$\begin{aligned} \tau \frac{dh_i^{\mathrm{HD}}(t)}{dt} = &-h_i^{\mathrm{HD}}(t) + \frac{\phi_0}{C^{\mathrm{HD}}} \sum_j (w_{ij} - w^{inh}) r_j^{\mathrm{HD}}(t) + I_i^V \\ &+ \phi_1 \left(\frac{1}{C^{\mathrm{HD}\times\mathrm{ID}}} \sum_{j,k} w_{ijk}^{\mathrm{ID}} r_j^{\mathrm{HD}} r_k^{\mathrm{ID}} \right). \end{aligned} \tag{16.9}$$

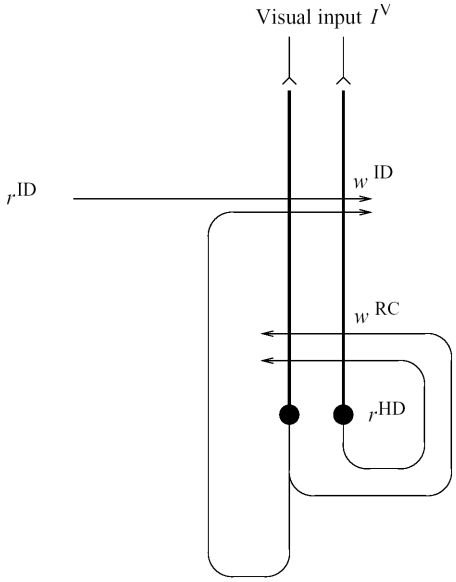

Figure 16.7

General network architecture for a one-dimensional continuous attractor model of head direction cells which can be updated by idiothetic inputs produced by head rotation cell firing r^{ID}. The head direction cell firing is r^{HD}, the continuous attractor synaptic weights are w^{RC}, the idiothetic synaptic weights are w^{ID}, and the external visual input is I^{V}.

Equation 16.9 is similar to equation 16.3, except for the last term, which introduces the effects of the idiothetic synaptic weights w^{ID}_{ijk}, which effectively specify that the current firing of head direction cell i, r^{HD}_i, must be updated by the previously learned combination of the particular head rotation now occurring indicated by r^{ID}_k, and the current head direction indicated by the firings of the other head direction cells r^{HD}_j indexed through j[2]. This makes it clear that the idiothetic synapses operate using combinations of inputs, in this case of two inputs. Neurons that sum the effects of such local products are termed Sigma-Pi neurons. Although such synapses are more complicated than the two-term synapses used throughout the rest of this book, such three-term synapses appear to be useful to solve the computational problem of updating representations based on idiothetic inputs in the way described. Synapses that operate according to Sigma-Pi rules might be implemented in the brain by a number of mechanisms described by [38] (Section 21.1.1), [36], and [101], including

[2]The term $\phi_1/C^{\mathrm{HD}\times\mathrm{ID}}$ is a scaling factor that reflects the number $C^{\mathrm{HD}\times\mathrm{ID}}$ of inputs to these synapses, and enables the overall magnitude of the idiothetic input to each head direction cell to remain approximately the same as the number of idiothetic connections received by each head direction cell is varied.

having two inputs close together on a thin dendrite, so that local synaptic interactions would be emphasized.

Simulations demonstrating the operation of this self-organizing learning to produce movement of the location being represented in a continuous attractor network were described by [101], and one example of the operation is shown in Figure 16.2.4.2. They also showed that, after training with just one value of the head rotation cell firing, the network showed the desirable property of moving the head direction being represented in the continuous attractor by an amount that was proportional to the value of the head rotation cell firing. [101] also describe a related model of the idiothetic cell update of the location represented in a continuous attractor, in which the rotation cell firing directly modulates in a multiplicative way the strength of the recurrent connections in the continuous attractor in such a way that clockwise rotation cells modulate the strength of the synaptic connections in the clockwise direction in the continuous attractor, and vice versa. It should be emphasized that although the cells are organized in Figure 16.2.4.2 according to the spatial position being represented, there is no need for cells in continuous attractors that represent nearby locations in the state space to be close together, as the distance in the state space between any two neurons is represented by the strength of the connection between them, not by where the neurons are physically located. This enables continuous attractor networks to represent spaces with arbitrary topologies, as the topology is represented in the connection strengths [101, 102, 104, 105]. Indeed, it is this that enables many different charts each with its own topology to be represented in a single continuous attractor network [8].

16.2.4.3 Continuous attractor networks in two or more dimensions

Some types of spatial representation used by the brain are of spaces that exist in two or more dimensions. Examples are the two- (or three-) dimensional space representing where one is looking at in a spatial scene. Another is the two- (or three-) dimensional space representing where one is located. It is possible to extend continuous attractor networks to operate in higher dimensional spaces than the one-dimensional spaces considered so far [112, 104]. Indeed, it is also possible to extend the analyses of how idiothetic inputs could be used to update two-dimensional state spaces, such as the locations represented by place cells in rats [104] and the location at which one is looking represented by primate spatial view cells [105, 102]. Interestingly, the number of terms in the synapses implementing idiothetic update do not need to increase beyond three (as in Sigma-Pi synapses) even when higher dimensional state spaces are being considered [104]. Also interestingly, a continuous attractor network can in fact represent the properties of very high dimensional spaces, because the properties of the spaces are captured by the connections between the neurons of the continuous attractor, and these connections are of course, as in the world of discrete attractor networks, capable of representing high dimensional spaces [104]. With these approaches, continuous attractor networks have been developed of the two-dimensional representation of rat hippocampal place cells with idiothetic update by movements in the environment [104], and of primate hippocampal spatial view

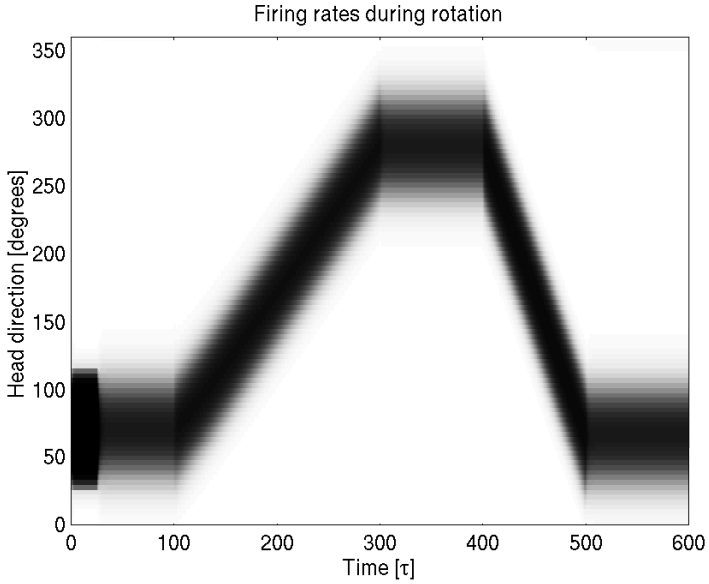

Figure 16.8

Idiothetic update of the location represented in a continuous attractor network. The firing rate of the cells with optima at different head directions (organized according to head direction on the ordinate) is shown by the blackness of the plot, as a function of time. The activity packet was initialized to a head direction of 75 degrees, and the packet was allowed to settle without visual input. For $t = 0$ to $t = 100$ there was no rotation cell input, and the activity packet in the continuous attractor remained stable at 75 degrees. For $t = 100$ to $t = 300$ the clockwise rotation cells were active with a firing rate of 0.15 to represent a moderate angular velocity, and the activity packet moved clockwise. For $t = 300$ to $t = 400$ there was no rotation cell firing, and the activity packet immediately stopped, and remained still. For $t = 400$ to $t = 500$ the anti-clockwise rotation cells had a high firing rate of 0.3 to represent a high velocity, and the activity packet moved anti-clockwise with a greater velocity. For $t = 500$ to $t = 600$ there was no rotation cell firing, and the activity packet immediately stopped.

cells with idiothetic update by eye and head movements [102, 105].

16.2.5 A unified theory of hippocampal memory: mixed continuous and dis-

crete attractor networks

If the hippocampus is to store and retrieve episodic memories, it may need to associate together patterns which have continuous spatial attributes, and other patterns which represent objects, which are discrete. To address this issue, we have now shown that attractor networks can store both continuous patterns and discrete patterns, and can thus be used to store for example the location in (continuous, physical) space where an object (a discrete item) is present (see Figure 16.4 and [88]). In this network, when events are stored that have both discrete (object) and continuous (spatial) aspects, then the whole place can be retrieved later by the object, and the object can be retrieved by using the place as a retrieval cue. Such networks are likely to be present in parts of the brain that receive and combine inputs both from systems that contain representations of continuous (physical) space, and from brain systems that contain representations of discrete objects, such as the inferior temporal visual cortex. One such brain system is the hippocampus, which appears to combine and store such representations in a mixed attractor network in the CA3 region, which thus is able to implement episodic memories which typically have a spatial component, for example where an item such as a key is located.

This network thus shows that in brain regions where the spatial and object processing streams are brought together, then a single network can represent and learn associations between both types of input. Indeed, in brain regions such as the hippocampal system, it is essential that the spatial and object processing streams are brought together in a single network, for it is only when both types of information are in the same network that spatial information can be retrieved from object information, and vice versa, which is a fundamental property of episodic memory. It may also be the case that in the prefrontal cortex, attractor networks can store both spatial and discrete (e.g. object-based) types of information in short term memory (see below).

16.2.6 The speed of operation of memory networks: the integrate-and-fire approach

Consider for example a real network whose operation has been described by an autoassociative formal model that acquires, with learning, a given attractor structure. How does the state of the network approach, in real time during a retrieval operation, one of those attractors? How long does it take? How does the amount of information that can be read off the network's activity evolve with time? Also, which of the potential steady states is indeed a stable state that can be reached asymptotically by the net? How is the stability of different states modulated by external agents? These are examples of dynamical properties, which to be studied require the use of models endowed with some dynamics. An appropriate such model is one which incorporates integrate-and-fire neurons.

The concept that attractor (autoassociation) networks can operate very rapidly if implemented with neurons that operate dynamically in continuous time is described by

[82] and [92]. The result described was that the principal factor affecting the speed of retrieval is the time constant of the synapses between the neurons that form the attractor ([7, 59, 92, 113]). This was shown analytically by [113], and described by [92] Appendix 5. If the (inactivation) time constant of AMPA synapses is taken as 10 ms, then the settling time for a single attractor network is approximately 15–17 ms [7, 59, 92]. A connected series of four such networks (representing for example four connected cortical areas) each involving recurrent (feedback) processing implemented by the recurrent collateral synaptic connections, takes approximately 4 x 17 ms to propagate from start to finish, retrieving information from each layer as the propagation proceeds [82, 59]. This speed of operation is sufficiently rapid that such attractor networks are biologically plausible [82, 92].

The way in which networks with continuous dynamics (such as networks made of real neurons in the brain, and networks modelled with integrate-and-fire neurons) can be conceptualized as settling so fast into their attractor states is that spontaneous activity in the network ensures that some neurons are close to their firing threshold when the retrieval cue is presented, so that the firing of these neurons is influenced within 1–2 ms by the retrieval cue. These neurons then influence other neurons within milliseconds (given the point that some other neurons will be close to threshold) through the modified recurrent collateral synapses that store the information. In this way, the neurons in networks with continuous dynamics can influence each other within a fraction of the synaptic time constant, and retrieval can be very rapid [92, 82].

16.3 Short term memory systems

16.3.1 Prefrontal cortex short term memory networks, and their relation to temporal and parietal perceptual networks

A common way that the brain uses to implement a short term memory is to maintain the firing of neurons during a short memory period after the end of a stimulus (see [24] and [92]). In the inferior temporal cortex this firing may be maintained for a few hundred ms even when the monkey is not performing a memory task [18, 89, 90, 91]. In more ventral temporal cortical areas such as the entorhinal cortex the firing may be maintained for longer periods in delayed match to sample tasks [109], and in the prefrontal cortex for even tens of seconds [23, 24]. In the dorsolateral and inferior convexity prefrontal cortex the firing of the neurons may be related to the memory of spatial responses or objects [30, 119] or both [63], and in the principal sulcus / arcuate sulcus region to the memory of places for eye movements [22] (see [82]). The firing may be maintained by the operation of associatively modified recurrent collateral connections between nearby pyramidal cells producing attractor states in

autoassociative networks (see [82]).
For the short term memory to be maintained during periods in which new stimuli are to be perceived, there must be separate networks for the perceptual and short term memory functions, and indeed two coupled networks, one in the inferior temporal visual cortex for perceptual functions, and another in the prefrontal cortex for maintaining the short term memory during intervening stimuli, provide a precise model of the interaction of perceptual and short term memory systems [67, 70] (see Figure 16.9). In particular, this model shows how a prefrontal cortex attractor (autoassociation) network could be triggered by a sample visual stimulus represented in the inferior temporal visual cortex in a delayed match to sample task, and could keep this attractor active during a memory interval in which intervening stimuli are shown. Then when the sample stimulus reappears in the task as a match stimulus, the inferior temporal cortex module showed a large response to the match stimulus, because it is activated both by the visual incoming match stimulus, and by the consistent backprojected memory of the sample stimulus still being represented in the prefrontal cortex memory module (see Figure 16.9). This computational model makes it clear that in order for ongoing perception to occur unhindered implemented by posterior cortex (parietal and temporal lobe) networks, there must be a separate set of modules that is capable of maintaining a representation over intervening stimuli. This is the fundamental understanding offered for the evolution and functions of the dorsolateral prefrontal cortex, and it is this ability to provide multiple separate short term attractor memories that provides we suggest the basis for its functions in planning. [67] and [70] performed analyses and simulations which showed that for working memory to be implemented in this way, the connections between the perceptual and the short term memory modules (see Figure 16.9) must be relatively weak. As a starting point, they used the neurophysiological data showing that in delayed match to sample tasks with intervening stimuli, the neuronal activity in the inferior temporal visual cortex (IT) is driven by each new incoming visual stimulus [64, 66], whereas in the prefrontal cortex, neurons start to fire when the sample stimulus is shown, and continue the firing that represents the sample stimulus even when the potential match stimuli are being shown [65]. The architecture studied by [70] was as shown in Figure 16.9, with both the intramodular (recurrent collateral) and the intermodular (forward IT to PF, and backward PF to IT) connections trained on the set of patterns with an associative synaptic modification rule. A crucial parameter is the strength of the intermodular connections, g, which indicates the relative strength of the intermodular to the intramodular connections. This parameter measures effectively the relative strengths of the currents injected into the neurons by the intermodular relative to the intra-modular connections, and the importance of setting this parameter to relatively weak values for useful interactions between coupled attractor networks was highlighted by [68] and [69] (see [82]). The patterns themselves were sets of random numbers, and the simulation utilized a dynamical approach with neurons with continuous (hyperbolic tangent) activation functions (see Section 16.3.2 and [5, 40, 41, 96]). The external current injected into IT by the incoming visual stimuli was sufficiently strong to trigger the IT module into a state representing the incoming stimulus. When the sample was shown, the initially silent PF module was

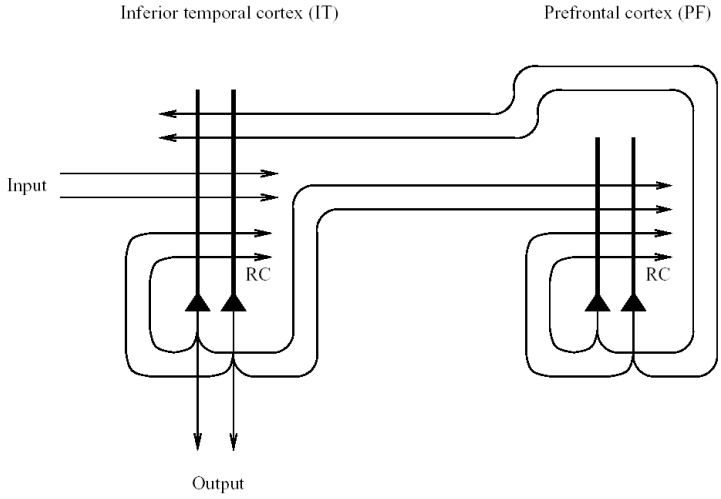

Figure 16.9

A short term memory autoassociation network in the prefrontal cortex could hold active a working memory representation by maintaining its firing in an attractor state. The prefrontal module would be loaded with the to-be-remembered stimulus by the posterior module (in the temporal or parietal cortex) in which the incoming stimuli are represented. Backprojections from the prefrontal short term memory module to the posterior module would enable the working memory to be unloaded, to for example influence on-going perception (see text). RC - recurrent collateral connections.

triggered into activity by the weak ($g > 0.002$) intermodular connections. The PF module remained firing to the sample stimulus even when IT was responding to potential match stimuli later in the trial, provided that g was less than 0.024, because then the intramodular recurrent connections could dominate the firing (see Figure 16.10). If g was higher than this, then the PF module was pushed out of the attractor state produced by the sample stimulus. The IT module responded to each incoming potentially matching stimulus provided that g was not greater than approximately 0.024. Moreover, this value of g was sufficiently large that a larger response of the IT module was found when the stimulus matched the sample stimulus (the match enhancement effect found neurophysiologically, and a mechanism by which the matching stimulus can be identified). This simple model thus shows that the operation of the prefrontal cortex in short term memory tasks such as delayed match to sample with intervening stimuli, and its relation to posterior perceptual networks, can be understood by the interaction of two weakly coupled attractor networks, as shown in Figs. 16.9 and 16.10.

The same network can also be used to illustrate the interaction between the prefrontal cortex short term memory system and the posterior (IT or PP) perceptual regions in

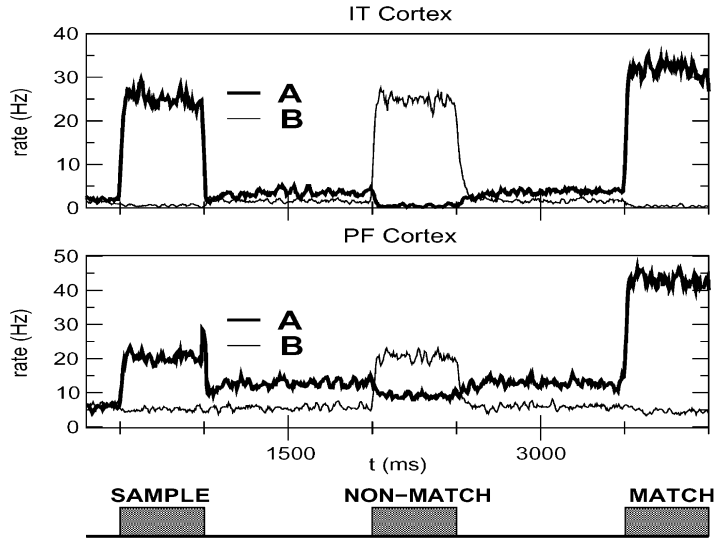

Figure 16.10

Interaction between the prefrontal cortex (PF) and the inferior temporal cortex (IT) in a delayed match to sample task with intervening stimuli with the architecture illustrated in Figure 16.9. Above: activity in the IT attractor module. Below: activity in the PF attractor module. The thick lines show the firing rates of the set of neurons with activity selective for the Sample stimulus (which is also shown as the Match stimulus, and is labelled A), and the thin lines the activity of the neurons with activity selective for the Non-Match stimulus, which is shown as an intervening stimulus between the Sample and Match stimulus and is labelled B. A trial is illustrated in which A is the Sample (and Match) stimulus. The prefrontal cortex module is pushed into an attractor state for the sample stimulus by the IT activity induced by the sample stimulus. Because of the weak coupling to the PF module from the IT module, the PF module remains in this Sample-related attractor state during the delay periods, and even while the IT module is responding to the non-match stimulus. The PF module remains in its Sample-related state even during the Non-Match stimulus because once a module is in an attractor state, it is relatively stable. When the Sample stimulus reappears as the Match stimulus, the PF module shows higher Sample stimulus-related firing, because the incoming input from IT is now adding to the activity in the PF attractor network. This in turn also produces a match enhancement effect in the IT neurons with Sample stimulus-related selectivity, because the backprojected activity from the PF module matches the incoming activity to the IT module. After Renart, Parga and Rolls, 2000 and Renart, Moreno, de la Rocha, Parga and Rolls, 2001.

visual search tasks, as illustrated in Figure 16.11.

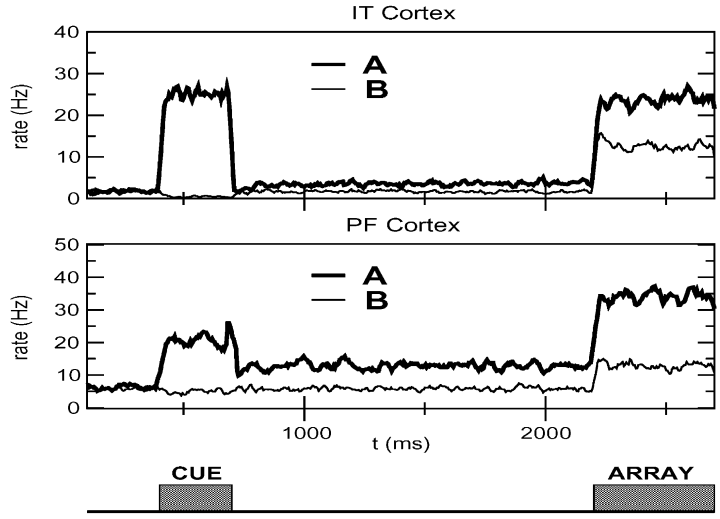

Figure 16.11

Interaction between the prefrontal cortex (PF) and the inferior temporal cortex (IT) in a visual search task with the architecture illustrated in Figure 16.9. Above: activity in the IT attractor module. Below: activity in the PF attractor module. The thick lines show the firing rates of the set of neurons with activity selective for search stimulus A, and the thin lines the activity of the neurons with activity selective for stimulus B. During the cue period either A or B is shown, to indicate to the monkey which stimulus to select when an array containing both A and B is shown after a delay period. The trial shown is for the case when A is the cue stimulus. When stimulus A is shown as a cue, then via the IT module, the PF module is pushed into an attractor state A, and the PF module remembers this state during the delay period. When the array A + B is shown later, there is more activity in the PF module for the neurons selective for A, because they have inputs both from the continuing attractor state held in the PF module and from the forward activity from the IT module which now contains both A and B. This PF firing to A in turn also produces greater firing of the population of IT neurons selective for A than in the IT neurons selective for B, because the IT neurons selective for A are receiving both A–related visual inputs, and A–related backprojected inputs from the PF module. After Renart, Parga and Rolls, 2000 and Renart, Moreno, de la Rocha, Parga and Rolls, 2001.

16.3.2 Computational details of the model of short term memory

The model network of [67] and [70] consists of a large number of (excitatory) neurons arranged in two modules with the architecture shown in Figure 16.9. Following

[5, 40], each neuron is assumed to be a dynamical element which transforms an incoming afferent current into an output spike rate according to a given transduction function. A given afferent current I_{ai} to neuron i $(i = 1, \ldots, N)$ in module a $(a = \mathbf{IT}, \mathbf{PF})$ decays with a characteristic time constant τ but increases proportionally to the spike rates of the rest of the neurons in the network (both from inside and outside its module) connected to it, the contribution of each presynaptic neuron, e.g. neuron j from module b, and in proportion to the synaptic efficacy J_{ij}^{ab} between the two[3]. This can be expressed through the following equation

$$\frac{dI_{ai}(t)}{dt} = -\frac{I_{ai}(t)}{\tau} + \sum_{b,j} J_{ij}^{(a,b)} \nu_{bj} + h_{ai}^{(\text{ext})} \quad . \tag{16.10}$$

An external current $h_{ai}^{(\text{ext})}$ from outside the network, representing the stimuli, can also be imposed on every neuron. Selective stimuli are modelled as proportional to the stored patterns, i.e. $h_{ai}^{\mu_{(\text{ext})}} = h_a \eta_{ai}^{\mu}$, where h_a is the intensity of the external current to module a.

The transduction function of the neurons transforming currents into rates was chosen as a threshold hyperbolic tangent of gain G and threshold θ. Thus, when the current is very large the firing rates saturate to an arbitrary value of 1.

The synaptic efficacies between the neurons of each module and between the neurons in different modules are respectively

$$J_{ij}^{(a,a)} = \frac{J_0}{f(1-f)N_t} \sum_{\mu=1}^{P} (\eta_{ai}^{\mu} - f)\,(\eta_{aj}^{\mu} - f) \qquad i \neq j \;\; ; \;\; a = \mathbf{IT}, \mathbf{PF} \tag{16.11}$$

$$J_{ij}^{(a,b)} = \frac{g}{f(1-f)N_t} \sum_{\mu=1}^{P} (\eta_{ai}^{\mu} - f)\,(\eta_{bj}^{\mu} - f) \qquad \forall \;\; i, j \;\; ; \;\; a \neq b \;\; . \tag{16.12}$$

The intra-modular connections are such that a number P of sparse independent configurations of neural activity are dynamically stable, constituting the possible sustained activity states in each module. This is expressed by saying that each module has learned P binary patterns $\{\eta_{ai}^{\mu} = 0, 1, \;\; \mu = 1, \ldots, P\}$, each of them signalling which neurons are active in each of the sustained activity configurations. Each variable η_{ai}^{μ} is allowed to take the values 1 and 0 with probabilities f and $(1 - f)$ respectively, independently across neurons and across patterns. The inter-modular connections reflect the temporal associations between the sustained activity states of each module. In this way, every stored pattern μ in the IT module has an associated pattern in the PF module which is labelled by the same index. The normalization constant $N_t = N(J_0 + g)$ was chosen so that the sum of the magnitudes of the inter- and the intra-modular connections remains constant and equal to 1 while their

[3] On this occasion we revert to the theoretical physicists' usual notation for synaptic weights or couplings, J_{ij}, from w_{ij}.

relative values are varied. When this constraint is imposed the strength of the connections can be expressed in terms of a single independent parameter g measuring the relative intensity of the inter- vs. the intra-modular connections (J_0 can be set equal to 1 everywhere).
Both modules implicitly include an inhibitory population of neurons receiving and sending signals to the excitatory neurons through uniform synapses. In this case the inhibitory population can be treated as a single inhibitory neuron with an activity dependent only on the mean activity of the excitatory population. We chose the transduction function of the inhibitory neuron to be linear with slope γ.
Since the number of neurons in a typical network one may be interested in is very large, e.g. $\sim 10^5 - 10^6$, the analytical treatment of the set of coupled differential equations (16.10) becomes untractable. On the other hand, when the number of neurons is large, a reliable description of the asymptotic solutions of these equations can be found using the techniques of statistical mechanics [40]. In this framework, instead of characterizing the states of the system by the state of every neuron, this characterization is performed in terms of *macroscopic* quantities called *order parameters* which measure and quantify some global properties of the network as a whole. The relevant order parameters appearing in the description of the system are the overlap of the state of each module with each of the stored patterns m_a^μ and the average activity of each module x_a, defined respectively as:

$$m_a^\mu = \frac{1}{\chi N} \ll \sum_i (\eta_{ai}^\mu - f)\nu_{ai} \gg_\eta \;\; ; \;\; x_a = \frac{1}{N} \ll \sum_i \nu_{ai} \gg_\eta \quad , \tag{16.13}$$

where the symbol $\ll \ldots \gg_\eta$ stands for an average over the stored patterns.
Using the free energy per neuron of the system at zero temperature $\mathcal{F}$ (which is not written explicitly to reduce the technicalities to a minimum), [70] and [67] modelled the experiments by giving the order parameters the following dynamics:

$$\tau \frac{\partial m_a^\mu}{\partial t} = -\frac{\partial \mathcal{F}}{\partial m_a^\mu} \quad ; \quad \tau \frac{\partial x_a}{\partial t} = -\frac{\partial \mathcal{F}}{\partial x_a} \quad . \tag{16.14}$$

These dynamics ensure that the stationary solutions, corresponding to the values of the order parameters at the attractors, correspond also to minima of the free energy, and that, as the system evolves, the free energy is always minimized through its gradient. The time constant of the macroscopical dynamics was chosen to be equal to the time constant of the individual neurons, which reflects the assumption that neurons operate in parallel. Equations (16.14) were solved by a simple discretizing procedure (first order Runge-Kutta method). An appropriate value for the time interval corresponding to one computer iteration was found to be $\tau/10$ and the time constant has been given the value $\tau = 10\ ms$.
Since not all neurons in the network receive the same inputs, not all of them behave in the same way, i.e. have the same firing rates. In fact, the neurons in each of the modules can be split into different sub-populations according to their state of activity in each of the stored patterns. The mean firing rate of the neurons in each sub-population depends on the particular state realized by the network (characterized

by the values of the order parameters). Associated with each pattern there are two large sub-populations denoted as foreground (all active neurons) and background (all inactive neurons) for that pattern. The overlap with a given pattern can be expressed as the difference between the mean firing rate of the neurons in its foreground and its background. The average was calculated over all other sub-populations to which each neuron in the foreground (background) belonged to, where the probability of a given sub-population is equal to the fraction of neurons in the module belonging to it (determined by the probability distribution of the stored patterns as given above). This partition of the neurons into sub-populations is appealing since, in neurophysiological experiments, cells are usually classified in terms of their response properties to a set of fixed stimuli, i.e. whether each stimulus is effective or ineffective in driving their response.

The modelling of the different experiments proceeded according to the macroscopic dynamics (16.14), where each stimulus was implemented as an extra current into free energy for a desired period of time.

Using this model, results of the type described in Section 16.3.1 were found [67, 70]. The paper by [67] extended the earlier findings of [70] to integrate-and-fire neurons, and it is results from the integrate-and-fire simulations that are shown in Figs. 16.10 and 16.11.

16.3.3 Computational necessity for a separate, prefrontal cortex, short term memory system

This approach emphasizes that in order to provide a good brain lesion test of prefrontal cortex short term memory functions, the task set should require a short term memory for stimuli over an interval in which other stimuli are being processed, because otherwise the posterior cortex perceptual modules could implement the short term memory function by their own recurrent collateral connections. This approach also emphasizes that there are many at least partially independent modules for short term memory functions in the prefrontal cortex (e.g. several modules for delayed saccades; one or more for delayed spatial (body) responses in the dorsolateral prefrontal cortex; one or more for remembering visual stimuli in the more ventral prefrontal cortex; and at least one in the left prefrontal cortex used for remembering the words produced in a verbal fluency task – see Section 10.3 of [92]).

This computational approach thus provides a clear understanding of why a separate (prefrontal) mechanism is needed for working memory functions, as elaborated in Section 16.3.1. It may also be commented that if a prefrontal cortex module is to control behaviour in a working memory task, then it must be capable of assuming some type of executive control. There may be no need to have a single central executive additional to the control that must be capable of being exerted by every short term memory module. This is in contrast to what has traditionally been assumed for the prefrontal cortex [98].

16.3.4 Role of prefrontal cortex short term memory systems in visual search and attention

The same model shown in Figure 16.9 can also be used to help understand the implementation of visual search tasks in the brain [70]. In such a visual search task, the target stimulus is made known beforehand, and inferior temporal cortex neurons then respond more when the search target (as compared to a different stimulus) appears in the receptive field of the IT neuron [16, 15]. The model shows that this could be implemented by the same system of weakly coupled attractor networks in PF and IT shown in Figure 16.9 as follows. When the target stimulus is shown, it is loaded into the PF module from the IT module as described for the delayed match to sample task. Later, when the display appears with two or more stimuli present, there is an enhanced response to the target stimulus in the receptive field, because of the backprojected activity from PF to IT which adds to the firing being produced by the target stimulus itself [67, 70] (see Figure 16.11). The interacting spatial and object networks described by [82]) in Chapters 9–11, take this analysis one stage further, and show that once the PF–IT interaction has set up a greater response to the search target in IT, this enhanced response can in turn by backprojections to topologically mapped earlier cortical visual areas move the "attentional spotlight" to the place where the search target is located.

16.3.5 Synaptic modification is needed to set up but not to reuse short term memory systems

To set up a new short term memory attractor, synaptic modification is needed to form the new stable attractor. Once the attractor is set up, it may be used repeatedly when triggered by an appropriate cue to hold the short term memory state active by continued neuronal firing even without any further synaptic modification (see [37] and [82]). Thus manipulations that impair the long term potentiation of synapses (LTP) may impair the formation of new short term memory states, but not the use of previously learned short term memory states. [37] analyzed many studies of the effects of blockade of LTP in the hippocampus on spatial working memory tasks, and found evidence consistent with this prediction. Interestingly, it was found that if there was a large change in the delay interval over which the spatial information had to be remembered, then the task became susceptible, during the transition to the new delay interval, to the effects of blockade of LTP. The implication is that some new learning is required when the rat must learn the strategy of retaining information for longer periods when the retention interval is changed.

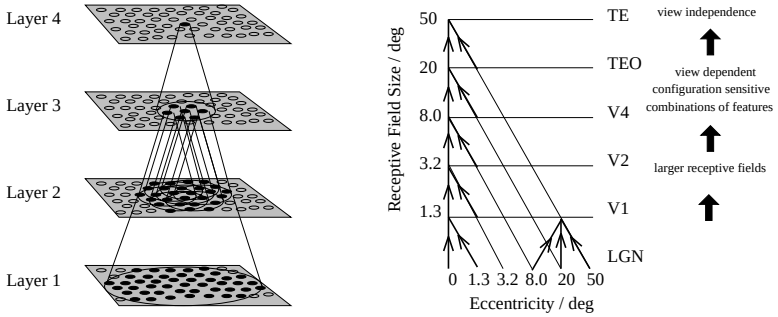

Figure 16.12

Convergence in the visual system. Right – as it occurs in the brain. V1: visual cortex area V1; TEO: posterior inferior temporal cortex; TE: inferior temporal cortex (IT). Left – as implemented in VisNet. Convergence through the network is designed to provide fourth layer neurons with information from across the entire input retina.

16.4 Invariant visual object recognition

[74] proposed a feature hierarchical model of ventral stream visual object processing from the primary visual cortex (V1), via V2 and V4 to the inferior temporal visual cortex which could learn to represent objects invariantly with respect to position on the retina, scale, rotation and view. The theory uses a short term ('trace') memory term in an associative learning rule to help capture the fact that the natural statistics of the visual world reflect the fact that the same object is likely to be present over short time periods, for example over 1 or 2 seconds during which an object is seen from different views. A model of the operation of the system has been implemented in a four-layer network, corresponding to brain areas V1, V2, V4 and inferior temporal visual cortex (IT), with convergence to each part of a layer from a small region of the preceding layer, and with local competition between the neurons within a layer implemented by local lateral inhibition [20, 82, 83, 117] (see Figure 16.12). During a learning phase each object is learned. This is done by training the connections between modules using a trace learning rule with the general form

$$\delta w_{ij} = \alpha \overline{y}_i{}^{\tau} x_j^{\tau} \tag{16.15}$$

where x_j^{τ} is the jth input to the neuron at time step τ, y_i is the output of the ith neuron, and w_{ij} is the jth weight on the ith neuron.
The trace $\overline{y}_i{}^{\tau}$ is updated according to

$$\overline{y}_i{}^{\tau} = (1-\eta) y_i{}^{\tau} + \eta \overline{y}_i{}^{\tau-1}. \tag{16.16}$$

The parameter $\eta \in [0, 1]$ controls the relative contributions to the trace $\overline{y}_i{}^{\tau}$ from the instantaneous firing rate $y_i{}^{\tau}$ at time step τ and the trace at the previous time step $\overline{y}_i{}^{\tau-1}$.

16.5 Visual stimulus–reward association, emotion, and motivation

Learning about which visual and other stimuli in the environment are rewarding, punishing, or neutral is crucial for survival. For example, it takes just one trial to learn if a seen object is hot when we touch it, and associating that visual stimulus with the pain may help us to avoid serious injury in the future. Similarly, if we are given a new food which has an excellent taste, we can learn in one trial to associate the sight of it with its taste, so that we can select it in future. In these examples, the previously neutral visual stimuli become conditioned reinforcers by their association with a primary (unlearned) reinforcer such as taste or pain. Our examples show that learning about which stimuli are rewards and punishments is very important in the control of motivational behaviour such as feeding and drinking, and in emotional behaviour such as fear and pleasure. The type of learning involved is pattern association, between the conditioned and the unconditioned stimulus. This type of learning provides a major example of how the visual representations provided by the inferior temporal visual cortex are used by the other parts of the brain [77, 80, 82]. In this Section we consider where in sensory processing this stimulus-reinforcement association learning occurs, which brain structures are involved in this type of learning, how the neuronal networks for pattern association learning may actually be implemented in these regions, and how the distributed representation about objects provided by the inferior temporal cortex output is suitable for this pattern association learning.

The crux of the answer to the last question is that the inferior temporal cortex representation is ideal for this pattern association learning because it is a transform-invariant representation of objects, and because the code can be read by a neuronal system which performs dot products using neuronal ensembles as inputs, which is precisely what pattern associators in the brain need, because they are implemented by neurons which perform as their generic computation a dot product of their inputs with their synaptic weight vectors (see [82] and [92]).

A schematic diagram summarizing some of the conclusions reached [77, 82, 92] is shown in Figure 16.13. The pathways are shown with more detail in Figure 16.14. The primate inferior temporal visual cortex provides a representation that is independent of reward or punishment, and is about objects. The utility of this is that the output of the inferior temporal visual cortex can be used for many memory and related functions (including episodic memory, short term memory, and reward/punishment memory) independently of whether the visual stimulus is currently rewarding or not. Thus we can learn about objects, and place them in short term memory, indepen-

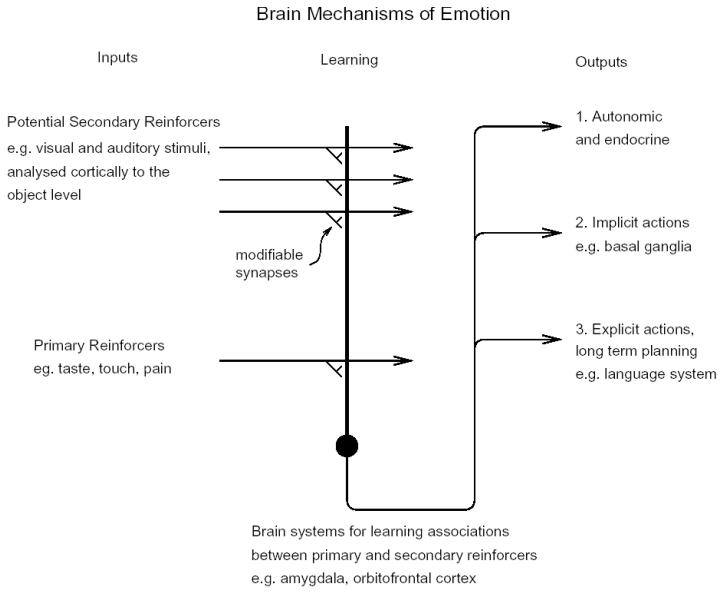

Figure 16.13

Schematic diagram showing the organization of brain networks involved in learning reinforcement associations of visual and auditory stimuli. The learning is implemented by pattern association networks in the amygdala and orbitofrontal cortex. The visual representation provided by the inferior temporal cortex is in an appropriate form for this pattern association learning, in that information about objects can be read from a population of IT neurons by dot-product neuronal operations.

dently of whether they are currently wanted or not. This is a key feature of brain design. The inferior temporal cortex then projects into two structures, the amygdala and orbitofrontal cortex, that contain representations of primary (unlearned) reinforcers such as taste and pain. These two brain regions then learn associations between visual and other previously neutral stimuli, and primary reinforcers [77], using what is highly likely to be a pattern association network, as illustrated in Figure 16.13. A difference between the primate amygdala and orbitofrontal cortex may be that the orbitofrontal cortex is set up to perform reversal of these associations very rapidly, in as little as one trial. Because the amygdala and orbitofrontal cortex represent primary reinforcers, and learn associations between these and neutral stimuli, they are key brain regions in emotions (which can be understood as states elicited by reinforcers, that is rewards and punishers), and in motivational states such as feeding and drinking [77].

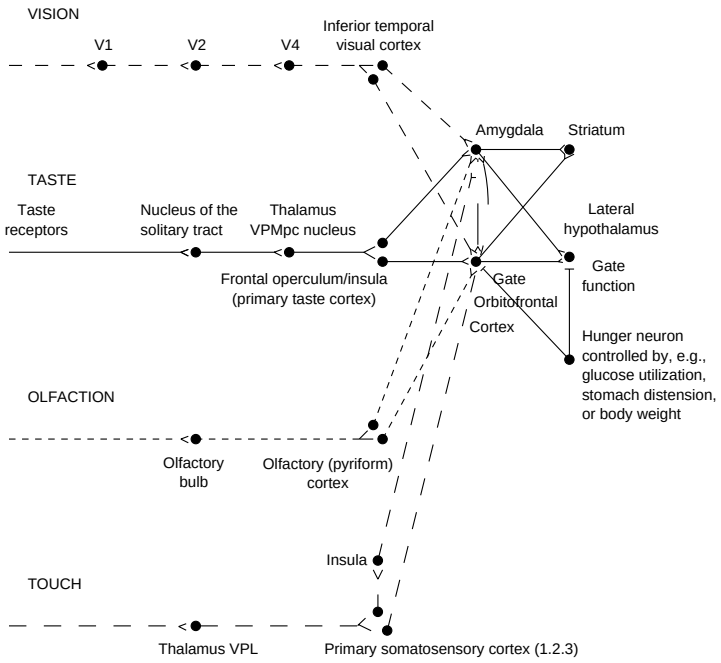

Figure 16.14

Diagrammatic representation of some of the connections described in this chapter. V1, striate visual cortex. V2 and V4, cortical visual areas. In primates, sensory analysis proceeds in the visual system as far as the inferior temporal cortex and the primary gustatory cortex; beyond these areas, in for example the amygdala and orbitofrontal cortex, the hedonic value of the stimuli, and whether they are reinforcing or are associated with reinforcement, is represented (see text).

16.6 Effects of mood on memory and visual processing

The current mood state can affect the cognitive evaluation of events or memories (see [9], [87]). An example is that when they are in a depressed mood, people tend to recall memories that were stored when they were depressed. The recall of depressing memories when depressed can have the effect of perpetuating the depression, and this may be a factor with relevance to the etiology and treatment of depression. A normal function of the effects of mood state on memory recall might be to facilitate continuity in the interpretation of the reinforcing value of events in the environment, or in the interpretation of an individual's behaviour by others, or simply to keep behaviour motivated to a particular goal. Another possibility is that the effects of mood on memory do not have adaptive value, but are a consequence of having a general

cortical architecture with backprojections. According to the latter hypothesis, the selection pressure is great for leaving the general architecture operational, rather than trying to find a genetic way to switch off backprojections just for the projections of mood systems back to perceptual systems (cf. [86]).

[87] (see also [75] and [77]) have developed a theory of how the effects of mood on memory and perception could be implemented in the brain. The architecture, shown in Figure 16.15, uses the massive backprojections from parts of the brain where mood is represented, such as the orbitofrontal cortex and amygdala to the cortical areas such as the inferior temporal visual cortex and hippocampus-related areas (labelled IT in Figure 16.15) that project into these mood-representing areas [2, 1]. The model uses an attractor in the mood module (labelled amygdala in Figure 16.15), which helps the mood to be an enduring state, and also an attractor in IT. The system is treated as a system of coupled attractors (see [82]), but with an odd twist: many different perceptual states are associated with any one mood state. Overall, there is a large number of perceptual / memory states, and only a few mood states, so that there is a many-to-one relation between perceptual / memory states and the associated mood states. The network displays the properties that one would expect (provided that the coupling parameters g between the attractors are weak). These include the ability of a perceptual input to trigger a mood state in the 'amygdala' module if there is not an existing mood, but greater difficulty to induce a new mood if there is already a strong mood attractor present; and the ability of the mood to affect via the backprojections which memories are triggered.

An interesting property which was revealed by the model is that because of the many-to-few mapping of perceptual to mood states, an effect of a mood was that it tended to make all the perceptual or memory states associated with a particular mood more similar then they would otherwise have been. The implication is that the coupling parameter g for the backprojections must be quite weak, as otherwise interference increases in the perceptual / memory module (IT in Figure 16.15).

Acknowledgments: This research was supported by Medical Research Council Programme Grant PG9826105, by the MRC Interdisciplinary Research Centre for Cognitive Neuroscience, and by the Human Frontier Science Program.

References

[1] Amaral, D. G., and Price, J. L. (1984), Amygdalo-cortical projections in the monkey (Macaca fascicularis), *Journal of Comparative Neurology*, **230**, 465–496.

[2] Amaral, D. G., Price, J. L., Pitkanen, A., and Carmichael, S. T. (1992), Anatomical organization of the primate amygdaloid complex, in Aggleton, J. P. (ed.) *The Amygdala*, Wiley-Liss: New York, 1–66.

[3] Amari, S. (1977), Dynamics of pattern formation in lateral-inhibition type neural fields, *Biological Cybernetics*, **27**, 77–87.

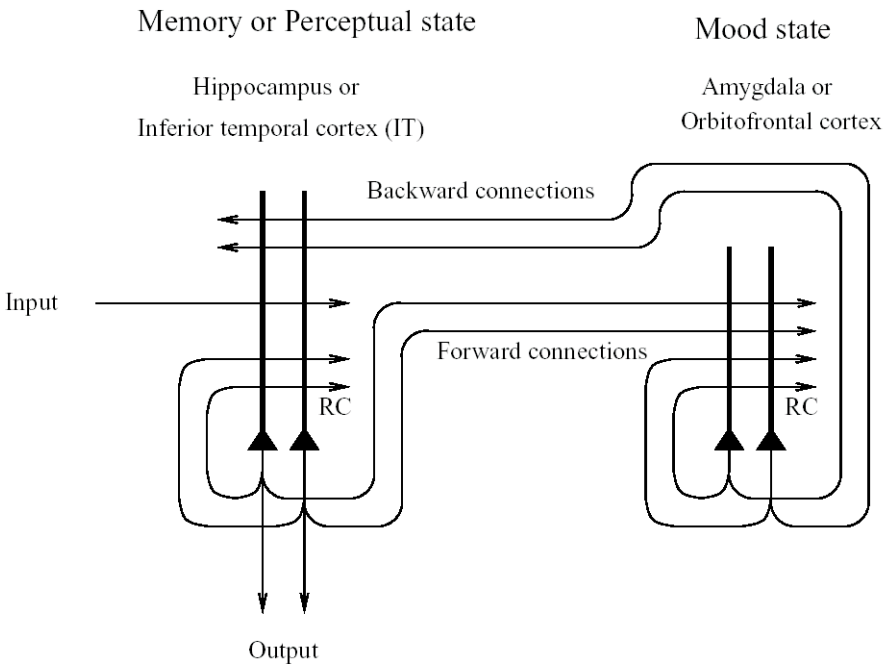

Figure 16.15

Architecture used to investigate how mood can affect perception and memory. The IT module represents brain areas such as the inferior temporal cortex involved in perception and hippocampus-related cortical areas that have forward connections to regions such as the amygdala and orbitofrontal cortex involved in mood. (After Rolls and Stringer (2001b)).

[4] Amit, D. J. (1989), *Modelling Brain Function*, Cambridge University Press: New York.

[5] Amit, D. J., and Tsodyks, M. V.(1991), Quantitative study of attractor neural network retrieving at low spike rates. I. Substrate – spikes, rates and neuronal gain, *Network*, **2**, 259–273.

[6] Andersen, R. A., Batista, A. P., Snyder, L. H., Buneo, C. A., and Cohen, Y. E. (2000), Programming to look and reach in the posterior parietal cortex, in Gazzaniga, M.S. (ed.) *The New Cognitive Neurosciences*, 2 ed., MIT Press: Cambridge, MA, 515–524.

[7] Battaglia, F., and Treves, A. (1998), Stable and rapid recurrent processing in realistic autoassociative memories, *Neural Computation* **10**, 431–450.

[8] Battaglia, F. P., and Treves, A. (1998), Attractor neural networks storing multiple space representations: A model for hippocampal place fields, *Physical*

Review E, **58**, 7738–7753.

[9] Blaney, P. H. (1986), Affect and memory: a review, *Psychological Bulletin*, **99**, 229–246.

[10] Burgess, N., and O'Keefe, J. (1996), Neuronal computations underlying the firing of place cells and their role in navigation, *Hippocampus*, **6**, 749–762.

[11] Burgess, N., Recce, M., and O'Keefe, J. (1994), A model of hippocampal function, *Neural Networks*, **7**, 1065–1081.

[12] Buckley, M. J., and Gaffan, D. (2000), The hippocampus, perirhinal cortex, and memory in the monkey, in Bolhuis, J. J. (ed.) *Brain, Perception, and Memory: Advances in Cognitive Neuroscience*, Oxford University Press: Oxford, 279–298.

[13] Cahusac, P. M. B., Rolls, E. T., Miyashita, Y., and Niki, H. (1993), Modification of the responses of hippocampal neurons in the monkey during the learning of a conditional spatial response task, *Hippocampus*, **3**, 29–42.

[14] Cassaday, H. J., and Rawlins, J. N. (1997), The hippocampus, objects, and their contexts, *Behavioural Neuroscience*, **111**, 1228–1244.

[15] Chelazzi, L., Duncan, J., Miller, E., and Desimone, R. (1998), Responses of neurons in inferior temporal cortex during memory-guuided visual search, *Journal of Neurophysiology*, **80**, 2918–2940.

[16] Chelazzi, L., Miller, E., Duncan, J., and Desimone, R. (1993), A neural basis for visual search in inferior temporal cortex, *Nature* (London), **363**, 345–347.

[17] de Araujo, I. E. T., Rolls, E. T., and Stringer, S. M. (2001), A view model which accounts for the response properties of hippocampal primate spatial view cells and rat place cells, *Hippocampus* , **11**, 699–706.

[18] Desimone, R. (1996), Neural mechanisms for visual memory and their role in attention, *Proceedings of the National Academy of Sciences USA*, **93**, 13494–13499.

[19] Eichenbaum, H. (1997), Declarative memory: insights from cognitive neurobiology, *Annual Review of Psychology*, **48**, 547–572.

[20] Elliffe, M. C. M., Rolls, E. T., and Stringer, S. M. (2001), Invariant recognition of feature combinations in the visual system, *Biological Cybernetics*, (in press).

[21] Epstein, R., and Kanwisher, N. (1998), A cortical representation of the local visual environment, *Nature*, **392**, 598–601.

[22] Funahashi, S., Bruce, C.J., and Goldman-Rakic, P.S. (1989), Mnemonic coding of visual space in monkey dorsolateral prefrontal cortex", *Journal of Neurophysiology*,**61**, 331–349.

[23] Fuster, J.M. (1997), *The Prefrontal Cortex*, 3rd, Raven Press: New York.

[24] Fuster, J.M. (2000), *Memory Systems in the Brain*, Raven Press: New York.

[25] Gaffan, D. (1994), Scene-specific memory for objects: a model of episodic memory impairment in monkeys with fornix transection, *Journal of Cognitive Neuroscience*, **6**, 305–320.

[26] Gaffan, D., and Harrison, S. (1989), A comparison of the effects of fornix section and sulcus principalis ablation upon spatial learning by monkeys, *Behavioural Brain Research*, **31**, 207–220.

[27] Gaffan, D., and Harrison, S. (1989), Place memory and scene memory: effects of fornix transection in the monkey, *Experimental Brain Research*, **74**, 202–212.

[28] Gaffan, D., and Saunders, R. C. (1985), Running recognition of configural stimuli by fornix transected monkeys, *Quarterly Journal of Experimental Psychology*, **37B**, 61–71.

[29] Georges-Francois, P., Rolls, E. T., and Robertson, R. G. (1999), Spatial view cells in the primate hippocampus: allocentric view not head direction or eye position or place, *Cerebral Cortex*, **9**, 197–212.

[30] Goldman-Rakic, P. S. (1996), The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive, *Philosophical Transactions of the Royal Society of London, Series B*, **351**, 1445–1453.

[31] Hasselmo, M. E., Schnell, E., and Barkai, E. (1995), Learning and recall at excitatory recurrent synapses and cholinergic modulation in hippocampal region CA3, *Journal of Neuroscience*, **15**, 5249–5262.

[32] Hertz, J., Krogh, A., and Palmer, R. G. (1991), *Introduction to the theory of neural computation*, Addison Wesley: Wokingham, U.K.

[33] Hopfield, J. J. (1982), Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences of the U.S.A.*, **79**, 2554–2558.

[34] Jackson, P. A., Kesner, R. P., and Amann, K. (1998), Memory for duration: role of hippocampus and medial prefrontal cortex, *Neurobiology of Learning and Memory*, **70**, 328–348.

[35] Jarrard, E. L. (1993), On the role of the hippocampus in learning and memory in the rat, *Behavioral and Neural Biology*, **60**, 9–26.

[36] Jonas, E. A., and Kaczmarek, L. K (1999), in Katz, P. S. (ed.) , *The inside story: subcellular mechanisms of neuromodulation*, Oxford University Press: New York 83–120.

[37] Kesner,R.P., and Rolls, E. T. (2001), Role of long term synaptic modification in short term memory, *Hippocampus*, **11**, 240-250.

[38] Koch, C. (1999), *Biophysics of Computation*, Oxford University Press: Oxford.

[39] Kubie, J. L., and Muller, R. U. (1991), Multiple representations in the hip-

pocampus, *Hippocampus*, **1**, 240-242.

[40] Kuhn, R. (1990),Statistical mechanics of neural networks near saturation, in Garrido, L. (ed.), *Statistical Mechanics of Neural Networks*, Springer-Verlag: Berlin.

[41] Kuhn, R., Bos, S., and van Hemmen, J. L. (1991), Statistical mechanics for networks of graded response neurons, *Physical Review A*, **243**, 2084–2087.

[42] Lassalle, J. M., Bataille, T., and Halley, H. (2000), Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the Morris navigation task, *Neurobiology of Learning and Memory*, **73**, 243–257.

[43] Markus, E. J., Qin, Y. L., Leonard, B. , Skaggs, W., McNaughton, B. L., and Barnes, C. A. (1995), Interactions between location and task affect the spatial and directional firing of hippocampal neurons, *Journal of Neuroscience*, **15**, 7079–7094.

[44] Marr, D. (1971), Simple memory: a theory for archicortex, *Philosophical Transactions of The Royal Society of London, Series B*, **262**, 23–81.

[45] Martin, S. J., Grimwood, P. D., and Morris, R. G. (2000), Synaptic plasticity and memory: an evaluation of the hypothesis, *Annual Review of Neuroscience*, **23**, 649–711.

[46] McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995), Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory, *Psychological Review*, **102**, 419–457.

[47] McNaughton, B. L., Barnes, C. A., and O'Keefe, J. (1983), The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats., *Experimental Brain Research*, **52**, 41–49.

[48] Miyashita, Y., Rolls, E. T., Cahusac, P. M. B. , Niki, H., and Feigenbaum, J. D. (1989), Activity of hippocampal neurons in the monkey related to a conditional spatial response task, *Journal of Neurophysiology*, **61**, 669–678.

[49] Muller, R. U., Kubie, J. L., Bostock, E. M., Taube, J. S., and Quirk, G. J. (1991), Spatial firing correlates of neurons in the hippocampal formation of freely moving rats, in Paillard, J. (ed.), *Brain and Space*, Oxford University Press: Oxford, 296–333.

[50] Muller, R. U., Ranck, J. B., and Taube, J. S. (1996), Head direction cells: properties and functional significance, *Current Opinion in Neurobiology*, **6**, 196–206.

[51] O'Keefe, J. (1979), A review of the hippocampal place cells, *Progress in Neurobiology*, **13**, 419–439.

[52] O'Keefe, J. (1984), Spatial memory within and without the hippocampal sys-

tem, in Seifert, W. (ed.) *Neurobiology of the Hippocampus*, Academic Press: London, 375–403.

[53] O'Keefe, J. (1990), A computational theory of the cognitive map, *Progress in Brain Research*, **83**, 301–312.

[54] O'Keefe, J. (1991), The hippocampal cognitive map and navigational strategies, in Paillard, J. (ed.), *Brain and Space*, Oxford University Press:Oxford, 273–295.

[55] O'Keefe, J., Burgess, N., Donnett, J. G., Jeffery, K. J., and Maguire, E. A. (1998), Place cells, navigational accuracy, and the human hippocampus,*Philosophical Transactions of the Royal Society, London [B]*, **353**, 1333–1340.

[56] O'Keefe, J., and Dostrovsky, J. (1971), The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat, *Brain Research*, **34** , 171–175.

[57] O'Keefe, J., and Nadel, L. (1978), *The Hippocampus as a Cognitive Map*, Clarendon Press: Oxford.

[58] O'Mara, S. M., Rolls, E. T., Berthoz, A., and Kesner, R. P. (1994), Neurons responding to whole-body motion in the primate hippocampus, *Journal of Neuroscience*, **14**, 6511–6523.

[59] Panzeri, S., Rolls, E. T., Battaglia, F., and Lavis, R. (2001), Speed of information retrieval in multilayer networks of integrate-and-fire neurons, *Network: Computation in Neural Systems*, **12**, 423–440.

[60] Parkinson, J. K., Murray, E. A., and Mishkin, M. (1988), A selective mnemonic role for the hippocampus in monkeys: memory for the location of objects, *Journal of Neuroscience*, **8**, 4059–4167.

[61] Petrides, M. (1985), Deficits on conditional associative-learning tasks after frontal- and temporal-lobe lesions in man, *Neuropsychologia*, **23**, 601–614.

[62] Ranck, Jr., J. B. (1985), Head direction cells in the deep cell layer of dorsolateral presubiculum in freely moving rats, in Buzsáki, G. and Vanderwolf, C. H. (eds.) *Electrical Activity of the Archicortex*, Akadémiai Kiadó: Budapest.

[63] Rao, S.C., Rainer, G., and Miller, E.K. (1997), Integration of what and where in the primate prefrontal cortex, *Science*, **276**, 821-824.

[64] Miller, E. K., and Desimone, R. (1994), Parallel neuronal mechanisms for short-term memory, *Science*, **263**, 520–522.

[65] Miller, E. K., Erickson, C., and Desimone, R. (1996), Neural mechanism of visual working memory in prefrontal cortex of the macaque, *Journal of Neuroscience*, **16**, 5154–5167.

[66] Miller, E. K., Li, L., and Desimone, R. (1993), Activity of neurons in anterior inferior temporal cortex during a short-term memory task, *Journal of Neuroscience*, **13**, 1460–1478.

[67] Renart, A., Moreno, R., de al Rocha, J., Parga, N., and Rolls, E. T. (2001), A model of the IT–PF network in object working memory which includes balanced persistent activity and tuned inhibition, *Neurocomputing*, **38–40**, 1525–1531.

[68] Renart, A., Parga, N., and Rolls, E. T. (1999), Backprojections in the cerebral cortex: implications for memory storage, *Neural Computation*, **11**, 1349–1388.

[69] Renart, A., Parga, N., and Rolls, E. T. (1999), Associative memory properties of multiple cortical modules, *Network*, **10**, 237–255.

[70] Renart, A., Parga, N.,and Rolls, E. T. (2000), A recurrent model of the interaction between the prefrontal cortex and inferior temporal cortex in delay memory tasks, in Solla, S.A. and Leen, T.K. and Mueller, K.-R. (eds.) *Advances in Neural Information Processing Systems*, MIT Press:Cambridge Mass, **12**, 171–177.

[71] Robertson, R. G., Rolls, E. T., and Georges-François, P. (1998), Spatial view cells in the primate hippocampus: Effects of removal of view details, *Journal of Neurophysiology*, **79**, 1145–1156.

[72] Robertson, R. G., Rolls, E. T., Georges-François, P., and Panzeri, S. (1999), Head direction cells in the primate pre-subiculum, *Hippocampus*, **9**, 206–219.

[73] Rolls, E. T. (1987), Information representation, processing and storage in the brain: analysis at the single neuron level, in *The Neural and Molecular Bases of Learning*, Changeux, J.-P. and Konishi, M. (ed.), Wiley: Chichester, 503–540.

[74] Rolls, E. T. (1992), Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas, *Philosophical Transactions of the Royal Society*, **335**, 11–21.

[75] Rolls, E. T. (1989), Functions of neuronal networks in the hippocampus and neocortex in memory, in Byrne, J.H. and Berry, W.O. (eds.), *Neural Models of Plasticity: Experimental and Theoretical Approaches*, Academic Press: San Diego, 240–265.

[76] Rolls, E. T. (1996), A theory of hippocampal function in memory, *Hippocampus*, **6**, 601–620.

[77] Rolls, E. T. (1999), The Brain and Emotion, Oxford University Press:Oxford.

[78] Rolls, E. T. (1999), Spatial view cells and the representation of place in the primate hippocampus, *Hippocampus*, **9**, 467–480.

[79] Rolls, E. T. (1999), The representation of space in the primate hippocampus, and its role in memory, *The Hippocampal and Parietal Foundations of Spatial Cognition*, in Burgess, N. and Jeffrey, K.J. and O'Keefe, J. (eds.), Oxford University Press: Oxford, 320–344.

[80] Rolls, E. T. (2000), Memory systems in the brain, *Annual Review of Psychology*, **51**, 599–630.

[81] Rolls, E. T. (2000), Hippocampo-cortical and cortico-cortical backprojections, *Hippocampus*, **10**, 380–388.

[82] Rolls, E. T., and Deco, G. (2002), *Computational Neuroscience of Vision*, Oxford University Press: Oxford.

[83] Rolls, E. T., and Milward, T. (2000), A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures, *Neural Computation*, **12**, 2547–2572.

[84] Rolls, E. T., Miyashita, Y., Cahusac, P. M. B., Kesner, R. P., Niki, H., Feigenbaum, J., and Bach, L. (1989), Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown, *Journal of Neuroscience*, **9**, 1835–1845.

[85] Rolls, E. T., Robertson, R. G., and Georges-François, P. (1997), Spatial view cells in the primate hippocampus, *European Journal of Neuroscience*, **9**, 1789–1794.

[86] Rolls, E. T., and Stringer, S. M. (2000), On the design of neural networks in the brain by genetic evolution, *Progress in Neurobiology*, **61**, 557–579.

[87] Rolls, E. T., and Stringer, S. M. (2001), A model of the interaction between mood and memory, *Network: Computation in Neural Systems*, **12**, 89-109.

[88] Rolls, E. T., Stringer, S. M., and Trappenberg, T. P. (2002), A unified model of spatial and episodic memory, *Proceedings of The Royal Society B*, **269**, 1087–1093.

[89] Rolls, E. T., and Tovee, M. J. (1994), Processing speed in the cerebral cortex and the neurophysiology of visual masking, *Proceedings of the Royal Society, B*, **257**, 9–15.

[90] Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., and Azzopardi, P. (1994), The responses of neurons in the temporal cortex of primates, and face identification and detection, *Experimental Brain Research*, **101**, 474–484.

[91] Rolls, E. T., Tovee, M. J., and Panzeri, S. (1999), The neurophysiology of backward visual masking: information analysis, *Journal of Cognitive Neuroscience*, **11**, 335–346.

[92] Rolls, E. T., and Treves, A. (1998), *Neural Networks and Brain Function*, Oxford University Press:Oxford.

[93] Rolls, E. T., Treves, A., Robertson, R. G., Georges-François, P., and Panzeri, S. (1998), Information about spatial view in an ensemble of primate hippocampal cells, *Journal of Neurophysiology*, **79**, 1797–1813.

[94] Rupniak, N. M. J., and Gaffan, D. (1987), Monkey hippocampus and learning about spatially directed movements, *Journal of Neuroscience*, **7**, 2331–2337.

[95] Samsonovich, A., and McNaughton, B.L. (1997), Path integration and cogni-

tive mapping in a continuous attractor neural network model, *Journal of Neuroscience*,

[96] Shiino, M., and Fukai, T. (1990), Replica-symmetric theory of the nonlinear analogue neural networks, *Journal of Physics A: Math. Gen.*, **23**, L1009–L1017.

[97] Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S., and McNaughton, B. L. (1995), A model of the neural basis of the rat's sense of direction, in Tesauro, G., Touretzky, D. S., and Leen, T. K.(eds.) *Advances in Neural Information Processing Systems*, vol. 7, 173–180, MIT Press: Cambridge, Massachusetts. **17**, 5900–5920.

[98] Shallice, T., and Burgess, P. (1996), The domain of supervisory processes and temporal organization of behaviour, *Philosophical Transactions of the Royal Society of London. Series B Biological Sciences*, **351**, 1405–1411.

[99] Smith, M. L., and Milner, B. (1981), The role of the right hippocampus in the recall of spatial location, *Neuropsychologia*, **19**, 781–793.

[100] Squire, L. R., and Knowlton, B. J. (2000) The medial temporal lobe, the hippocampus, and the memory systems of the brain, in *The New Cognitive Neurosciences*, Gazzaniga, M.S. (ed.), 765–779, 2nd, MIT Press: Cambridge, MA.

[101] Stringer, S. M., Trappenberg, T. P., Rolls, E. T., and de Araujo, I. E. T. (2002), Self-organizing continuous attractor networks and path integration: One-dimensional models of head direction cells, *"Network: Computation in Neural Systems*, **13**, 217–242.

[102] Stringer, S. M., and Rolls, E. T. (2002), Self-organizing continuous attractor network models of spatial view cells for an agent that is able to move freely through different locations, (submitted).

[103] Stringer, S. M., and Rolls, E. T. (2002), Hierarchical dynamical models of motor function, (submitted).

[104] Stringer, S. M., Rolls, E.T., Trappenberg, T. P., and de Araujo, I. E. T. (2002), Self-organizing continuous attractor networks and path integration: Two-dimensional models of place cells, *"Network: Computation in Neural Systems*, (in press).

[105] Stringer, S. M., Rolls, E. T., and Trappenberg, T. P. (2002), Self-organizing continuous attractor network models of hippocampal spatial view cells (in press).

[106] Stringer, S. M., and Rolls, E. T. (2002), Hierarchical dynamical models of motor function (in press).

[107] Suzuki, W. A., and Amaral, D. G. (1994), Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents, *Journal of Comparative Neurology*, **350**, 497–533.

[108] Suzuki, W. A., and Amaral, D. G. (1994), Topographic organization of the

reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices, *Journal of Neuroscience*, **14**, 1856–1877.

[109] Suzuki, W. A., Miller, E. K., and Desimone, R. (1997), Object and place memory in the macaque entorhinal cortex, *Journal of Neurophysiology*, **78**, 1062–1081.

[110] Taube, J. S., Goodridge, J. P., Golob, E. G., Dudchenko, P. A., and Stackman, R. W. (1996), *Processing the head direction signal: a review and commentary*, *Brain Research Bulletin*, **40**, 477–486.

[111] Taube, J. S., Muller, R. U., and Ranck, Jr., J. B. (1990), Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis, *Journal of Neuroscience* , **10**, 420–435.

[112] Taylor, J. G. (1999), *Neural 'bubble' dynamics in two dimensions: foundations*, *Biological Cybernetics*, **80**, 393–409.

[113] Treves, A. (1993), Mean-field analysis of neuronal spike dynamics Quantitative estimate of the information relayed by the Schaffer collaterals, *Network*, **4**, 259–284.

[114] Treves, A., and Rolls, E. T. (1991), What determines the capacity of autoassociative memories in the brain?, *Network*, **2**, 371–397.

[115] Treves, A., and Rolls, E. T. (1992), Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network, *Hippocampus*, **2**, 189–199.

[116] Treves, A., and Rolls, E. T. (1994), A computational analysis of the role of the hippocampus in memory, *Hippocampus*, **4**, 374–391.

[117] Wallis, G., and Rolls, E. T. (1997), Invariant face and object recognition in the visual system, *Progress in Neurobiology*, **51**, 167–194.

[118] Wilson, M. A., and McNaughton, B. L. (1993) Dynamics of the hippocampal ensemble code for space, *Science*, **261**, 1055-1058.

[119] Wilson, F. A. W., O'Sclaidhe, S. P., and Goldman-Rakic, P. S. (1993), Dissociation of object and spatial processing domains in primate prefrontal cortex, *Science*, **260**, 1955–1958.

[120] Zhang, K. (1996), Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory, *Journal of Neuroscience*, **16**, 2112–2126.

[121] Zhang, W. ,and Dietterich, T. G. (1996), High-performance job-shop scheduling with a time-delay TD(λ) Network, in Touretzky, D. S., Mozer, M. C., and Hasselmo, M. E. (eds.) Advances in Neural Information Processing Systems 8, Cambridge MA: MIT Press, 1024–1030.

[122] Zola-Morgan, S., Squire, L. R., Amaral, D. G., and Suzuki, W. A. (1989), Lesions of perirhinal and parahippocampal cortex that spare the amygdala and

hippocampal formation produce severe memory impairment, *Journal of Neuroscience*, **9**, 4355–4370.

[123] Zola-Morgan, S., Squire, L. R., and Ramus, S. J. (1994), Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal lobe memory system, *Hippocampus*, **4**, 483–494.