

Hippocampo-Cortical and Cortico-Cortical Backprojections

Edmund T. Rolls*

Department of Experimental Psychology, University of Oxford, Oxford, UK

ABSTRACT: First, the information represented in the primate hippocampus, and what is computed by the primate hippocampus, are considered. Then a theory is described of how the information represented in the hippocampus is able to influence the cerebral cortex by a hierarchy of hippocampo-cortical and cortico-cortical backprojection stages. The recalled backprojected information in the cerebral neocortex could then be used by the neocortex as part of memory recall, including that required in spatial working memory; to influence the processing that each cortical stage performs based on its forward inputs; to influence the formation of long-term memories; and/or in the selection of appropriate actions. *Hippocampus* 2000;10:380–388. © 2000 Wiley-Liss, Inc.

KEY WORDS: autoassociation; recall; episodic memory; cortical backprojections; attractor networks

THE INFORMATION REPRESENTED IN THE PRIMATE HIPPOCAMPUS

The systems-level neurophysiology of the hippocampus shows what information could be stored or processed by the hippocampus. To understand how the hippocampus works, it is not sufficient to state just that it can store information: one needs to know what information. The systems-level neurophysiology of the primate hippocampus has been reviewed in more detail by Rolls (1999), and only a brief summary is provided here.

The primate hippocampus contains spatial cells that respond when the monkey looks at a certain part of space, e.g., at one quadrant of a video monitor while the monkey is performing an object-place memory task in which he must remember where on the monitor he has seen particular images (Rolls et al., 1989). Approximately 9% of hippocampal neurons have such spatial view fields, and about 2.4% combine information about position in space with information about the object that is in that position in space (Rolls et al., 1989). The latter point shows that information from very different parts of the cerebral cortex (parietal for spatial information, and inferior temporal for visual information about objects) is brought together onto single neurons in the primate hippocampus. The representation of space is for the majority of hippocampal neurons in allocentric, not egocentric coordinates (Feigenbaum and Rolls, 1991; Rolls and O'Mara, 1995; Rolls et al., 1998; Georges-François et al., 1999).

These "spatial view" neurons, now analyzed in the actively locomoting monkey, are different from place cells, in that their activity is dependent not on the place where the monkey is, but on the place where the monkey is looking in space (Rolls et al., 1997). It can be shown with the monkey stationary that these cells respond when the monkey's eye position results in his looking at a particular part of space (Georges-François et al., 1999). Some of these spatial view cells respond when the view details are obscured by curtains or darkness, when the monkey's eyes look towards the spatial view field of the neuron (Robertson et al., 1999). In this situation, it is suggested that the neurons are responding to the recalled spatial

INTRODUCTION

Information reaches the primate hippocampus via a number of cortical stages, from the inferior temporal visual cortex to the perirhinal cortex, and from the parietal cortex to the parahippocampal cortex, which in turn projects to the entorhinal cortex, which then projects to the hippocampus (see Fig. 1). These are typical forward projections, in that they tend to project to the next stage from pyramidal cells in layers 2 and 3. Once information has reached the hippocampus, it must be capable of leaving if it is to be useful in any way. There are two main exit routes. One is via the fimbria/fornix to the anterior thalamus and mammillary bodies. The second is from CA1 via the entorhinal cortex and backprojections to the areas that originally sent inputs to the hippocampus. These are backprojections, in that in the cortical stages, the projections are mainly from cortical pyramidal cells in layers 5 and 6, and end on dendrites in layers 1 and 2 of the preceding cortical stage. The aim of this paper is to consider the functions of the hippocampo-cortical backprojections, and of cortico-cortical backprojections in general.

Grant sponsor: Medical Research Council; Grant number: PG98 26105; Grant sponsor: Human Frontier Science Program.

*Correspondence to: Professor E.T. Rolls, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK. E-mail: Edmund.Rolls@psy.ox.ac.uk.

Accepted for publication 1 May 2000

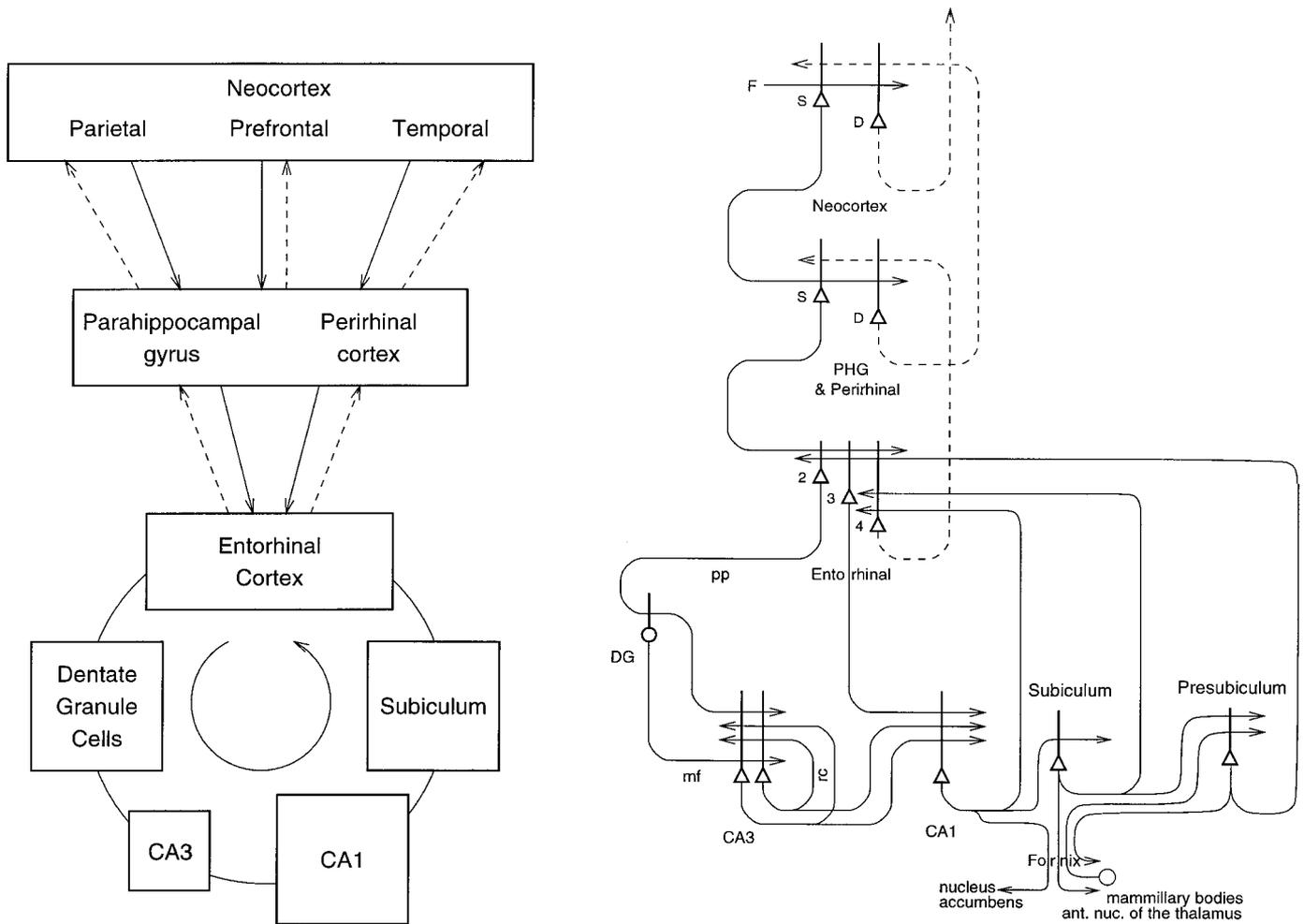


FIGURE 1. Forward connections (solid lines) from areas of cerebral association neocortex via the parahippocampal gyrus and perirhinal cortex, and entorhinal cortex, to the hippocampus; and backprojections (dashed lines) via the hippocampal CA1 pyramidal cells, subiculum, and parahippocampal gyrus to 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. 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 fibers; PHG, parahippocampal gyrus and perirhinal cortex; pp, perforant path; rc, recurrent collateral 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. Thick lines above cell bodies represent dendrites.

view, with the recall triggered by the partial idiothetic information provided about the spatial view by eye and head position information, which these experiments clearly do show influence primate hippocampal neurons. Many of these “spatial view” cells are hippocampal pyramidal cells, with very low spontaneous firing rates and low peak firing rates, and they implement a sparse representation. Although there is some evidence for place cells in the hippocampus of the monkey driving a cab (Ono et al., 1993), such place cells are not at all obvious in the actively locomoting monkey in a situation in which place cells would be found in rats (Rolls, 1999).

These findings indicate that, related to the highly developed visual system of primates, the spatial information represented in the primate hippocampus is mainly about space “out there,” rather than the place where the animal is, as in the rat. It seems to be a major use of the primate hippocampus to associate spatial locations

with what is there, and there is no need for the primate to visit the place for the cells to respond. Simply looking at the place and seeing an object, or person, in that place is perfectly adequate for primates, including humans, to form object-place memories, which allow them to later recall, for example, where the object or person was seen. It is for this type of spatial memory that it is suggested the representation of space just described in the primate hippocampus is used. The hippocampus would associate together the activity of one population of neurons representing spatial position “out there” with other neurons providing information about the object seen.

Other neurons in the primate hippocampus respond to combinations of visual object and spatial response information in associative learning tasks in which conditional spatial responses must be learned to visual images (Miyashita et al., 1989; Cahusac et al., 1993). The presence of these neurons provides additional evidence

that information represented in different parts of the cerebral cortex (visual temporal cortex and parietal cortex) is brought together on the same neurons in the hippocampus, in tasks in which those types of information must be rapidly associated. Other primate hippocampal neurons respond to (idiothetic) whole-body motion (O'Mara et al., 1994). These neurons represent information that it would be necessary to store in order to remember recent body movements made in short range navigation (so that one could perform path integration and for example return to the starting place); and to determine whether one was facing towards a particular location in space, i.e., they would provide an important input for "spatial view" neurons. Indeed, when spatial view neurons respond in the dark as a monkey is rotated towards a spatial view (Robertson et al., 1998), their activity may be triggered by whole-body motion cell input into the CA3 autoassociative network. Other neurons that may help with this process are head direction cells, which have been found in primates in the primate presubiculum (Robertson et al., 1999).

COMPUTATIONAL THEORIES OF THE HIPPOCAMPUS

One function performed by the primate hippocampus may be the association to form a memory of the activity of one population of neurons representing spatial position "out there" with other neurons providing information about the object seen.

Part of a theory of how this is performed is that the CA3 neurons provide an autoassociative network appropriate for rapidly (in one trial) learning such associations (Treves and Rolls, 1994; Rolls, 1989, 1996b). The theory is also compatible with the existence of place cells in rats, which might be formed by associating together a particular combination of spatial cues to define a place; and which would be useful in a memory system, for associating objects with the place where the rat found them. Indeed, it is suggested that part of the difference between rat place cells and primate spatial view cells is that with a wide field of view (on the order of 270°) in the rat, the association of widely spaced visual environmental cues might result in the formation of place cells, whereas in primates, the association of closely spaced visual cues given the much smaller field of view could result in spatial view cells (Stringer et al., 2000).

Under the hypothesis that the CA3 forms an autoassociative memory, the theory developed for attractor networks (Amit, 1989) with sparse representations and incomplete connectivity (Treves, 1990; Treves and Rolls, 1991) can be applied, and it can be shown that the maximum number of memory CA3 firing patterns p_{\max} that can be (individually) retrieved is proportional to the number C^{RC} of (associatively) modifiable recurrent collateral (RC) synapses per cell, by a factor that increases roughly with the inverse of the sparseness a of the neuronal representation. The sparseness is defined as

$$a = \left(\sum_{i=1,n} r_i/n \right)^2 / \sum_{i=1,n} (r_i^2/n) \quad (1)$$

where r_i is the firing rate to the i 'th stimulus in the set of n stimuli. The sparseness ranges from $1/n$, when the cell responds to only one stimulus, to a maximal value of 1.0, attained when the cell responds with the same rate to all stimuli. Approximately,

$$p_{\max} \approx \frac{C^{RC}}{a \ln(1/a)} k \quad (2)$$

where k is a factor that depends weakly on the detailed structure of the rate distribution, on the connectivity pattern, etc., but is roughly on the order of 0.2–0.3 (Treves and Rolls, 1991). For example, for $C^{RC} = 12,000$ and $a = 0.02$ (realistic estimates for the rat), p_{\max} is calculated to be approximately 36,000. Simulations that are fully consistent with the analytic theory are provided by Rolls et al. (1997), and by a fuller computational model (Rolls, 1995).

Another hypothesis is that the CA3 cells might provide a continuous attractor, i.e., an attractor network in which the synaptic strengths are a function of the distance between positions in a continuum (Samsonovich and McNaughton, 1997). The synaptic strengths in such a network might be formed by Hebbian associative learning, based on the overlapping firing rate response profiles of nearby hippocampal place (Samsonovich and McNaughton, 1997) or spatial view (Stringer et al., 2000) cells. The position in such a network would be represented by a group of firing cells (or activity packet), and the activity packet can move continuously from position to position in the continuous attractor. It has also been shown how multiple charts of place cells different in different environments could be stored in the CA3 network using a continuous attractor (cf. Battaglia and Treves, 1998). The hypothesis here for primates is that such a continuous attractor for spatial views might be updated in the hippocampus by idiothetic signals such as head direction, eye position, and whole-body motion in a way analogous to that suggested for the rat (Samsonovich and McNaughton, 1997). However, in the rat the proposed mechanism involves what is effectively a lookup table, with a whole set of neurons responding to all possible combinations of whole-body motion velocity and head direction (Samsonovich and McNaughton, 1997). In contrast, we have developed a model which involves a simple self-organizing learning rule with a memory trace of previous neuronal activity which can learn how to move the continuous attractor representing one spatial view (or place) to another, using idiothetic cues such as head direction and whole body motion (Stringer et al., 2000).

The computational differences between these theories are interesting. In the autoassociation theory, discrete representations, e.g., of spatial view and object, are associated together, and there are discrete memory states (minima in the energy landscape) to which the attractor settles to retrieve the memory. In contrast, with the continuous spatial attractor model, the energy landscape is flat, and the problem is how to move the firing rate activity packet from one location to another, using idiothetic cues. When the continuous attractor is accessed by visual cues, the neurons which represent a given spatial view can be accessed much more simply by associative lookup as in discrete attractor models, e.g., in the way described by Treves and Rolls (1992). One solution currently being explored is that the CA3 system may be kept stable in whatever position it is

currently firing (if there is no idiothetic input), thus performing a short-term memory of spatial location which could still operate in the dark, by virtue of being an amalgam of the two types of attractor network. This would allow for some bumps in the energy landscape for stability produced by virtue of discrete associations to visual cues or incomplete connectivity, but nevertheless sufficient continuity so that idiothetic inputs can move the activity pattern to another position (Stringer et al., 2000).

BACKPROJECTIONS TO THE NEOCORTEX

Given that some information is represented by the firing of hippocampal neurons, I now consider how that information could influence the neocortex via backprojections.

It is suggested that the modifiable connections from the CA3 neurons to the CA1 neurons allow the information present in CA3 to be produced efficiently in CA1 (Treves, 1995; Schultz and Rolls, 1999). The representation may be in a different form from that which is needed in CA3 for the recurrent collaterals to function correctly (see Rolls, 1996b). The CA1 neurons would then activate, via their termination in the deep layers of the entorhinal cortex, at least the pyramidal cells in the deep layers of the entorhinal cortex (see Fig. 1). These neurons would then, by virtue of their backprojections to the parts of cerebral cortex that originally provided the inputs to the hippocampus, terminate in the superficial layers of those neocortical areas, where synapses would be made onto the distal parts of the dendrites of the cortical pyramidal cells (see Rolls, 1989). The areas of cerebral neocortex in which this recall would be produced could include multimodal cortical areas (e.g., the cortex in the superior temporal sulcus which receives inputs from temporal, parietal, and occipital cortical areas), and also areas of unimodal association cortex (e.g., the inferior temporal visual cortex).

The hypothesis of the architecture with which this would be achieved is shown in Figure 1. The feedforward connections from association areas of the cerebral neocortex (solid lines in Fig. 1) show major convergence as information is passed to CA3, with the CA3 autoassociation network having the smallest number of neurons at any stage of the processing. The backprojections allow for divergence back to neocortical areas. The way in which I suggest that the backprojection synapses are set up to have the appropriate strengths for recall is as follows (see also Rolls, 1989). During the setting up of a new input to the hippocampus (perhaps for episodic memory), there would be strong feedforward activity progressing towards the hippocampus. During the episode, the CA3 synapses would be modified, and via the CA1 neurons and the subiculum, a pattern of activity would be produced on the backprojecting synapses to the entorhinal cortex. Here the backprojecting synapses from active backprojection axons onto pyramidal cells being activated by the forward inputs to entorhinal cortex would be associatively modified. A similar process would be implemented at preceding stages of the neocortex, i.e., in the parahippocampal gyrus/ perirhinal cortex stage, and in association cortical areas, as shown in Figure 1. The timing of the backprojecting activity would be

sufficiently rapid for this, in that, for example, inferior temporal cortex neurons become activated by visual stimuli with latencies of 90–110 ms and may continue firing for several hundred milliseconds (Rolls, 1992, 2000a,b); and hippocampal pyramidal cells are activated in visual object-and-place and conditional spatial response tasks with latencies of 120–180 ms (Rolls et al., 1989; Miyashita et al., 1989). Thus backprojected activity from the hippocampus might be expected to reach association cortical areas such as the inferior temporal visual cortex within 60–100 ms of the onset of their firing, and there would be a several hundred millisecond period in which there would be conjunctive feedforward activation present, with simultaneous backprojected signals in the association cortex.

During recall implemented by backprojections, the backprojection connections onto the distal synapses of cortical pyramidal cells would be helped in their efficiency in activating the pyramidal cells by virtue of two factors. The first is that with no forward input to the neocortical pyramidal cells, there would be little shunting of the effects received at the distal dendrites by the more proximal effects on the dendrite normally produced by the forward synapses. Further, without strong forward activation of the pyramidal cells, there would not be very strong feedback and feedforward inhibition via GABA cells, so that there would not be a further major loss of signal due to (shunting) inhibition on the cell body and (subtractive) inhibition on the dendrite. (The converse of this is that when forward inputs are present, as during normal processing of the environment rather than during recall, the forward inputs would, appropriately, dominate the activity of the pyramidal cells, which would be only influenced, not determined, by the backprojecting inputs; see Rolls, 1989.)

The synapses receiving the backprojections would have to be Hebb-modifiable, as suggested by Rolls (1989). This would solve the dead-dressing problem, i.e., the problem of how the hippocampus is able to bring into activity during recall just those cortical pyramidal cells that were active when the memory was originally being stored. The solution hypothesized (Rolls, 1989) arises because modification occurs during learning of the synapses from active backprojecting neurons from the hippocampal system onto the dendrites of only those neocortical pyramidal cells active at the time of learning. Without this modifiability of cortical backprojections during learning, it is difficult to see how exactly the correct cortical pyramidal cells active during the original learning experience would be activated during recall. Consistent with this hypothesis (Rolls, 1989), there are NMDA receptors present especially in superficial layers of the cerebral cortex (Monaghan and Cotman, 1985), implying Hebb-like learning just where the backprojecting axons make synapses with the apical dendrites of cortical pyramidal cells.

If the backprojection synapses are associatively modifiable, we may consider the duration of the period for which their synaptic modification should persist. What would be optimal would be to arrange for the associative modification of the backprojecting synapses to remain for as long as any associatively modified synapses remain in the hippocampus. This suggests that a similar mechanism for the associative modification would be appropriate both within the hippocampus and for that of at least one stage of the backprojecting synapses. It is suggested that the presence of high

concentrations of NMDA synapses in the distal parts of the dendrites of neocortical pyramidal cells and within the hippocampus may reflect the similarity of the synaptic modification processes in these two regions (cf. Kirkwood et al., 1993). It is noted that it would be appropriate to have this similarity of time course for at least one stage in the series of backprojecting stages from the CA3 region to the neocortex. Such stages might include the CA1 region, subiculum, entorhinal cortex, and perhaps the parahippocampal gyrus. However, from the multimodal cortex (e.g., the parahippocampal gyrus) back to earlier cortical stages, it might be desirable for the backprojecting synapses to persist for a long period, so that some types of recall and top-down processing (see Rolls, 1989) mediated by the operation of neocortico-neocortical backprojecting synapses could be stable.

An alternative hypothesis to that above is that rapid modifiability of backprojection synapses would be required only at the beginning of the backprojecting stream. Relatively fixed associations from higher to earlier neocortical stages would serve to activate the correct neurons at earlier cortical stages during recall. For example, there might be rapid modifiability from CA3 to CA1 neurons, but relatively fixed connections from there back (McClelland et al., 1995). For such a scheme to work, one would need to produce a theory not only of the formation of semantic memories in the neocortex, but also of how the operations performed according to that theory would lead to recall by setting up appropriately the backprojecting synapses.

BACKPROJECTIONS TO THE NEOCORTEX: QUANTITATIVE ASPECTS

How many backprojecting fibers does one need to synapse onto any given neocortical pyramidal cell, in order to implement the mechanism outlined above? Consider a polysynaptic sequence of backprojecting stages, from hippocampus to neocortex, as a string of simple (hetero-) associative memories in which, at each stage, the input connections are those coming from the previous stage (closer to the hippocampus). Implicit in this framework is the assumption that the synapses at each stage are modifiable according to some Hebbian associative plasticity rule. A plausible requirement for a successful hippocampo-directed recall operation is that the signal generated from the hippocampally retrieved pattern of activity, and carried backwards towards the neocortex, remain undegraded when compared to the noise due, at each stage, to the interference effects caused by the concurrent storage of other patterns of activity on the same backprojecting synaptic systems. That requirement is equivalent to that used in deriving the storage capacity of such a series of heteroassociative memories, and it was shown in Treves and Rolls (1991) that the maximum number of independently generated activity patterns that can be retrieved is given, essentially, by the same formula as (2) above:

$$p \approx \frac{C}{a \ln(1/a)} k' \quad (2')$$

where, however, a is now the sparseness of the representation at any given stage, and C is the average number of (back-)projections each cell of that stage receives from cells of the previous one (k' is a similar, slowly varying factor to that introduced above). If one assumes that a certain number of different attractor states can be determined by the CA3 recurrent collateral axons, then in order for all those states to be retrievable via the backprojecting synapses, there would need to be the same order of backprojection connections to each cell in the backprojecting pathway as there are recurrent collateral connections per neuron in the hippocampus (but dependent to some extent on the sparseness of the representation in the CA3 neurons and in the backprojecting pathways; see Treves and Rolls, 1994).

One consequence is that one could not possibly rely on a set of monosynaptic CA3-to-neocortex backprojections. This would imply that, to make a sufficient number of synapses on each of the vast number of neocortical cells, each cell in CA3 has to generate a disproportionate number of synapses (i.e., the number of backprojections received by a final cortical target cell times the ratio between the number of neocortical and of CA3 cells). The required divergence can be kept within reasonable limits only by assuming that the backprojecting system is polysynaptic, provided that the number of cells involved grows gradually at each stage, from CA3 back to neocortical association areas (cf. Fig. 1).

This theory provides what may be a fundamental insight into why there are approximately as many backprojections as forward-projecting neurons (between connected cortical areas, and between the hippocampus and neocortex). However, not all cortical backprojections might reflect backprojection signals originating from the hippocampus, for there are backprojections which might be considered to originate in the amygdala (see Amaral et al., 1992) or in multimodal cortical areas (e.g., allowing for recall of a visual image by an auditory stimulus with which it has been regularly associated). In this situation, one may consider whether the backprojections from any one of these systems would be sufficiently numerous to produce recall. One factor which may help here is that when recall is being produced by backprojections, it may be assisted by the local recurrent collaterals between nearby (~ 1 mm) pyramidal cells which are a feature of neocortical connectivity. These would tend to complete a partial neocortical representation being recalled by the backprojections into a complete recalled pattern.

There is evidence that the rat hippocampus is involved in some short term spatial working memory tasks (Kesner and Rolls, 2000), and when it is, the retrieval of information from the hippocampus may involve the types of process described here.

THE DYNAMICS OF RECURRENT NETWORKS, AND OF MULTISTAGE BACKPROJECTION NETWORKS

The analysis described above, of the capacity of a recurrent network such as the CA3 and of the backprojection pathways, was for steady-state conditions of the firing rates of neurons. The question arises of how quickly these networks would settle into their

final state. With reference to the CA3 network, how long does it take before a pattern of activity, originally evoked in CA3 by afferent inputs, becomes influenced by the activation of recurrent collaterals? In a more general context, recurrent collaterals between the pyramidal cells are an important feature of the connectivity of the cerebral neocortex. How long would it take these collaterals to contribute fully to the activity of cortical cells? If these settling processes took on the order of hundreds of milliseconds, they would be much too slow to contribute usefully to cortical activity, whether in the hippocampus or the neocortex (Rolls, 1992; Rolls and Treves, 1998).

A partial answer to this question can be inferred from a recent theoretical development based on the analysis of the collective dynamic properties of realistically modelled neuronal units (Treves, 1993). The method incorporates the biophysical properties of real cell membranes, and considers the dynamics of a network of integrate-and-fire neurons, laterally connected through realistically modelled synapses. The analysis indicates that the model network will attain a stable distribution of firing rates over time scales determined essentially by synaptic and intrinsic conductance inactivation times. Some of these (e.g., the conductance time constants associated with excitatory synapses between pyramidal cells) are very short, less than 10 ms, implying that the activation of recurrent collaterals between pyramidal cells will contribute to determine the overall firing pattern within a period of a very few tens of milliseconds (see Treves, 1993). In a simulation of such a system, we have obtained evidence for rapid recall, within 30–50 ms, even from a partial cue (Treves et al., 1997), and this was further addressed in a model with 10-compartment neurons to allow the shunting inhibition which is important for stability to be investigated (Battaglia and Treves, 1998). With respect to the CA3 network, the indication is thus that retrieval would be rapid, indeed fast enough for it to be biologically plausible.

I suggest that the dynamics of a multilayer backprojection network would also be very rapid, for similar reasons to the rapid settling of an attractor network that is implemented with neurons with continuous dynamics. In particular, provided that the neurons have spontaneous activity, some neurons will be close to their firing threshold when the retrieval process starts, and they will be triggered into activity very rapidly, within 1–3 ms. These neurons, in communicating to other neurons in the backprojecting pathways, will cause the firing of some of them to be influenced rapidly, and this will occur through the modified synapses in the backprojecting pathway. This should lead to continuous and rapid retrieval across even a multistage backprojection system. In a simulation of this with integrate-and-fire neurons, we have found that in the forward direction processing takes place in a few milliseconds per stage when there is no local recurrent processing within each stage, and takes only approximately 15 ms per stage of cortical processing when recurrent collateral feedback within each stage is required (Panzeri et al., 2000). This latter value matches rather well the time delays of cortical visual information processing from V1 to V2 to V4 to the inferior temporal visual cortex, a process that takes on the order of 50–70 ms (see Rolls and Treves, 1998; Panzeri et al., 2000). I conjecture that the backprojection time is very similar, i.e., a few milliseconds per stage for useful information

retrieval when the backprojection process does not require recurrent collateral feedback processing within each cortical stage, and approximately 17 ms per stage when it does. This is now being tested.

PERIRHINAL CORTEX BACKPROJECTIONS TO THE INFERIOR TEMPORAL VISUAL CORTEX

The perirhinal cortex may implement functions important for learning the view invariant representations of inferior temporal cortex neurons. The evidence for this is not only that inferior temporal cortex neurons have view invariant representations of objects (Booth and Rolls, 1998) and faces (Hasselmo et al., 1989; Rolls, 2000b), but also that perirhinal cortex lesions may impair the learning of new view-invariant representations of objects, but not the use of previously learned view invariant representations expressed by the activity of inferior temporal cortex neurons (Buckley et al., 1998).

One way in which the perirhinal cortex could help the inferior temporal cortex to learn view-invariant responses is by providing a short-term memory sufficiently long to bridge the gap between different views of a given object. It has been suggested that providing a memory trace in this way would enable a Hebbian synaptic modification to learn the correct associations between different views (Rolls, 1992). Such trace rule learning can implement translation, size, and view-invariant learning in a multistage hierarchical feedforward network (Wallis and Rolls, 1997; Rolls and Milward, 2000; Parga and Rolls, 1998; Rolls, 2000b). The perirhinal backprojections could contribute, using backprojections to the activity of inferior temporal cortex neurons to enable the latter to maintain their firing rates over the relevant time period of up to 1 s that typically in the real world separates different views of the same object. A similar process might help form neurons in the perirhinal cortex that can respond to the same pair of stimuli if they are closely and regularly paired together in temporal proximity.

BACKPROJECTIONS AND SHORT-TERM MEMORY

A common way that the brain implements a short-term memory is to maintain the firing of neurons during a short memory period after the end of a stimulus. In the inferior temporal cortex this firing may be maintained for a few hundred milliseconds, even when the monkey is not performing a memory task (Rolls and Tovee, 1994; Rolls et al., 1994, 1999; cf. Desimone, 1996). 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 (Suzuki et al., 1997), and in the prefrontal cortex for even tens of seconds (see Fuster, 1997). In the dorsolateral and inferior convexity prefrontal cortex the firing of the neurons may

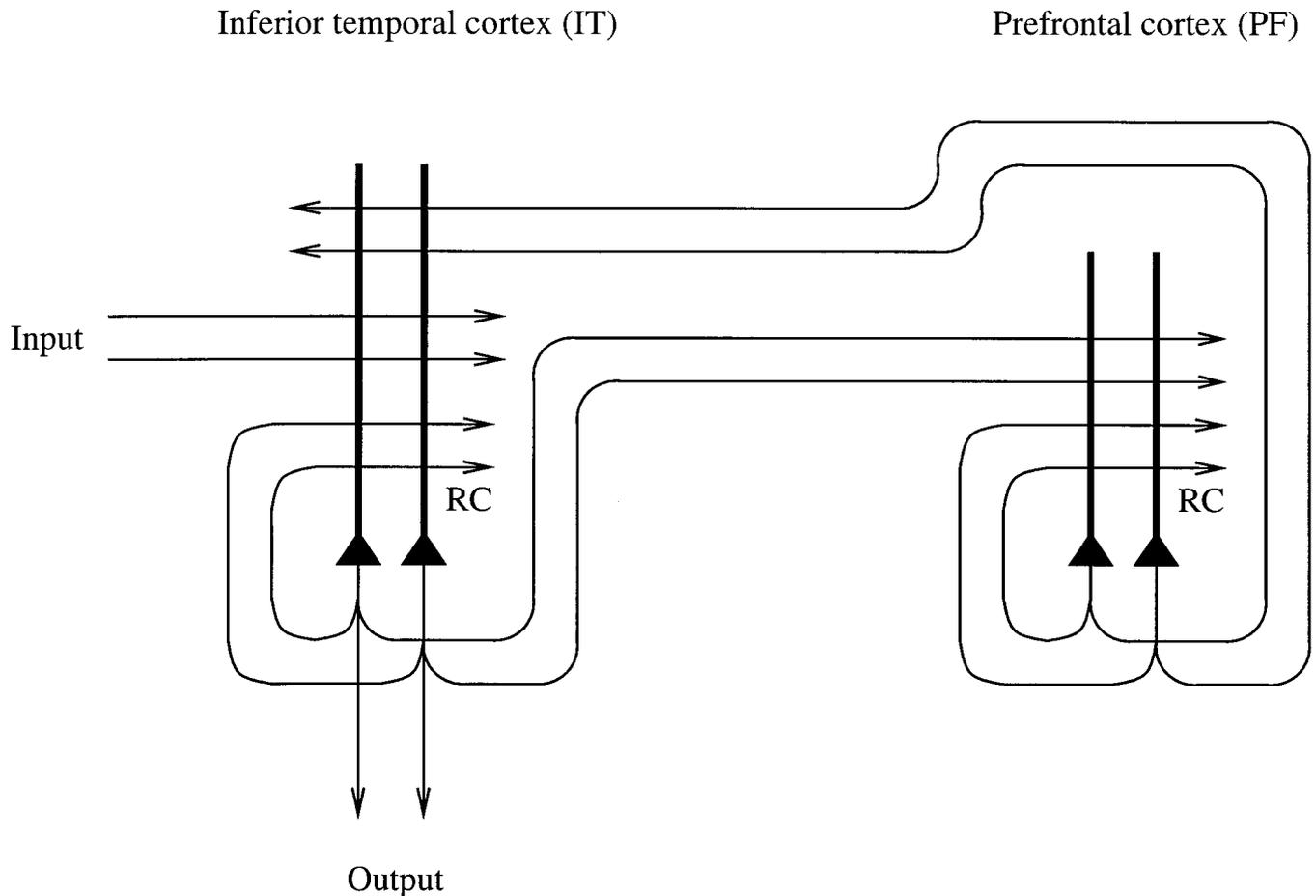


FIGURE 2. Two coupled attractor networks linked to provide a perceptual module (IT, in the inferior temporal/perirhinal cortex region) and a short-term memory module (PF, in the prefrontal cortex). The associative synaptic forward and backward (PF to IT) connections in the model must be much weaker than the intramodule recurrent collateral associative connections, in order for the system to allow interactions between the modules, and top-down modulatory influences, including effects such as the match enhancement effect in a delayed match to sample task with intervening stimuli, and the McGurk effect.

be related to the memory of spatial responses or objects (Goldman-Rakic, 1996, Wilson et al., 1993) or both (Rao et al., 1997), and in the principal sulcus/frontal eye field/arcuate sulcus region to the memory of directions for eye movements (Funahashi et al., 1989). The firing may be maintained by the operation of associatively modified recurrent collateral connections between nearby pyramidal cells, producing attractor states in autoassociative networks (Amit, 1995, Rolls and Treves, 1998). 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 (Renart et al., 1999c).

In the prefrontal cortex-inferior temporal (or perirhinal) cortex model analyzed by Renart et al. (2000b), each region operates as an attractor network, and the two attractor nets are coupled by forward associatively modifiable connections to the prefrontal cortex,

and by associatively modifiable backprojections to the inferior temporal cortex (Fig. 2). One effect that can be understood with such a model is the delayed match to sample with intervening stimuli task used by Miller and Desimone (1994). In this task a sample stimulus is shown, and then several intervening stimuli before the stimulus matching the sample is shown, when a behavioral response must be made. The model shows that, provided that the connections between the inferior temporal/perirhinal cortex (IT) perceptual model and the prefrontal (PF) short-term memory attractor networks are weak relative to the internal connectivity (the ratio must be on the order of 0.01 in the model), the sample stimulus produces activity in the IT and thence the PF module; the next and different (nonmatch) stimulus activates a different attractor in IT (because the forward connections to IT are not very weak), but this does not disturb the PF attractor (because it has weak inputs from IT), nor does the PF attractor for the stimulus disrupt the IT module (because the PF to IT connections are weak); yet when the match stimulus drives the IT module, an enhanced response is obtained (because the IT module now has consistent forward input and a backprojected signal from the PF

module). This model helps to show why separate networks are needed for perception and for some types of short-term memory, and makes the neurophysiologically interesting point that the backprojections between cortical modules need for at least some functions to be much weaker than the internal recurrent collateral connections within a module.

BACKPROJECTIONS BETWEEN CORTICAL MODULES

Renart et al. (1999a,b) also showed quantitatively how backprojections can influence earlier cortical stages. The model consists of two (Renart et al., 1999a) or three (Renart et al., 1999b) attractor networks coupled in the forward and backward directions by associatively modifiable synapses. Many properties of these coupled systems have been analyzed, including the small increase in memory capacity when corresponding memories are stored in the two modules; the presence of global attractor states when the two modules act as a network storing corresponding patterns in the different networks; the dependence of this global attractor scenario on the strength of the crosscoupling between the modules; the retrieval in one of the modules by an input applied only to another module; the effect of a nonnoisy input to one module, cleaning up the effects of a noisy input to another; and the influence of one module in causing another module to enter particular states, as in the McGurk effect, when what is seen can influence what is heard. The neurophysiologically interesting point is again made that if the two networks are not to collapse functionally into a single network, then the backprojections between cortical modules need to be much weaker than the internal recurrent collateral connections within a module. If this is the case, then interesting effects of backprojection top-down effects can be seen.

CONCLUSIONS

In this paper, a number of recent neurophysiological and theoretical investigations have been brought together to provide an indication not only of the processes that are performed within the primate hippocampus, but also of how hippocampo-cortical and cortico-cortical backprojections may operate.

Acknowledgments

Different parts of the research described here were supported by Medical Research Council grant PG98 26105, and by a Human Frontier Science Program grant. The author performed the experimental and theoretical work which is incorporated in the ideas presented here on the hippocampus with many colleagues, including Alessandro Treves, Robert Robertson, Pierre Georges-François, Ray Kesner, Shane O'Mara, Conrado Perez-Vicente, Martin

Simmen, and Simon Stringer, and their contributions are sincerely acknowledged.

REFERENCES

- Amaral DG, Price JL, Pitkanen A, Carmichael ST. 1992. Anatomical organization of the primate amygdaloid complex. In: Aggleton JP, editor. *The Amygdala*. New York: Wiley-Liss. p 1–66.
- Amit DJ. 1989. *Modelling brain function*. New York: Cambridge University Press.
- Amit DJ. 1995. The Hebbian paradigm reintegrated: local reverberations as internal representations. *Behav Brain Sci* 18:617–657.
- Battaglia FP, Treves A. 1998. Attractor neural networks storing multiple space representations: a model for hippocampal place fields. *Phys Rev [E]* 58:7738–7753.
- Booth MCA, Rolls ET. 1998. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex* 8:510–523.
- Buckley MJ, Booth MCA, Rolls ET, Gaffan D. 1998. Selective visual perception deficits following perirhinal cortex ablation in the macaque. *Soc Neurosci Abstr* 24:18.
- Cahusac PMB, Rolls ET, Miyashita Y, 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.
- Desimone R. 1996. Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci USA* 93:13494–13499.
- Feigenbaum JD, Rolls ET. 1991. Allocentric and egocentric spatial information processing in the hippocampal formation of the behaving primate. *Psychobiology* 19:21–40.
- Funahashi S, Bruce CJ, Goldman-Rakic PS. 1989. Mnemonic coding of visual space in monkey dorsolateral prefrontal cortex. *J Neurophysiol* 61:331–349.
- Fuster JM. 1997. *The prefrontal cortex*. 3rd ed. New York: Raven Press. p 333.
- Georges-François P, Rolls ET, Robertson RG. 1999. Spatial view cells in the primate hippocampus: allocentric view not head direction or eye position or place. *Cereb Cortex* 9:197–212.
- Goldman-Rakic PS. 1996. The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philos Trans R Soc Lond [Biol]* 351:1445–1453.
- Hasselmo ME, Rolls ET, Baylis GC, Nalwa V. 1989. Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Exp Brain Res* 75:417–429.
- Kesner RP, Rolls ET. 2000. Role of long term synaptic modification in short term memory.
- Kirkwood A, Duolek SM, Gold JT, Aizenman CD, Bear MF. 1993. Common forms of synaptic plasticity in the hippocampus and neocortex in vitro. *Science* 260:1518–1521.
- McClelland JL, McNaughton BL, O'Reilly RC. 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. *Psychol Rev* 102:419–457.
- Miller EK, Desimone R. 1994. Parallel mechanisms for short-term memory. *Science* 283:520–522.
- Miyashita Y, Rolls ET, Cahusac PMB, Niki H, Feigenbaum JD. 1989. Activity of hippocampal neurons in the monkey related to a conditional spatial response task. *J Neurophysiol* 61:669–678.
- Monaghan DT, Cotman CW. 1985. Distribution on N-methyl-D-aspartate-sensitive L-[3H]glutamate-binding sites in the rat brain. *J Neurosci* 5:2909–2919.
- O'Mara SM, Rolls ET, Berthoz A, Kesner RP. 1994. Neurons responding to whole-body motion in the primate hippocampus. *J Neurosci* 14:6511–6523.

- Ono T, Nakamura K, Nishijo H, Eifuku S. 1993. Monkey hippocampal neurons related to spatial and nonspatial functions. *J Neurophysiol* 70:1516–1529.
- Panzeri S, Rolls ET, Battaglia F, Lavis R. 2000. Speed of information retrieval in multilayer networks of integrate-and-fire neurons.
- Parga N, Rolls ET. 1998. Transform invariant recognition by association in a recurrent network. *Neural Comput* 10:1507–1525.
- Rao SC, Rainer G, Miller EK. 1997. Integration of what and where in the primate prefrontal cortex. *Science* 276:821–824.
- Renart A, Parga N, Rolls ET. 1999a. Backprojections in the cerebral cortex: implications for memory storage. *Neural Comput* 11:1349–1388.
- Renart A, Parga N, Rolls ET. 1999b. Associative memory properties of multiple cortical modules. *Network* 10:237–255.
- Renart A, Parga N, Rolls ET. 2000a. A recurrent model of the interaction between the prefrontal cortex and inferior temporal cortex in delay memory tasks. *Adv in Neural Inform Process Syst*. In press.
- Renart A, Parga N, Rolls ET. 2000b. The relation between perception and short term memory modelled by interacting recurrent network modules. *Cereb Cortex*.
- Robertson RG, Rolls ET, Georges-François P. 1998. Spatial view cells in the primate hippocampus: effects of removal of view details. *J Neurophysiol* 79:1145–1156.
- Robertson RG, Rolls ET, Georges-François P, Panzeri S. 1999. Head direction cells in the primate pre-subiculum. *Hippocampus* 9:206–219.
- Rolls ET. 1989. Functions of neuronal networks in the hippocampus and neocortex in memory. In: Byrne JH, Berry WO, editors. *Neural models of plasticity: experimental and theoretical approaches*. San Diego: Academic Press. p 240–265.
- Rolls ET. 1992. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philos Trans R Soc Lond [Biol]* 335:11–21.
- Rolls ET. 1995. A model of the operation of the hippocampus and entorhinal cortex in memory. *Int J Neural Syst [Suppl]* 6:51–70.
- Rolls ET. 1996a. Roles of long term potentiation and long term depression in neuronal network operations in the brain. In: Fazeli M, Collingridge G, editors. *Cortical plasticity: LTP and LTD*. Oxford: Bios. p 223–250.
- Rolls ET. 1996b. A theory of hippocampal function in memory. *Hippocampus* 6:601–620.
- Rolls ET. 1999. Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus* 9:467–480.
- Rolls ET. 2000a. Memory systems in the brain. *Annu Rev Psychol* 51:599–630.
- Rolls ET. 2000b. Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron*. In press.
- Rolls ET, 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 Comput* (in press).
- Rolls ET, O'Mara SM. 1995. View-responsive neurons in the primate hippocampal complex. *Hippocampus* 5:409–424.
- Rolls ET, Tovee MJ. 1994. Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc R Soc Lond [Biol]* 257:9–15.
- Rolls ET, Tovee MJ, Purcell DG, Stewart AL, Azzopardi P. 1994. The responses of neurons in the temporal cortex of primates, and face identification and detection. *Experimental Brain Research* 101:474–484.
- Rolls ET, Treves A. 1990. The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network* 1:407–421.
- Rolls ET, Treves A. 1998. *Neural networks and brain function*. Oxford: Oxford University Press.
- Rolls ET, Miyashita Y, Cahusac PMB, Kesner RP, Niki H, Feigenbaum J, Bach L. 1989. Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. *J Neurosci* 9:1835–1845.
- Rolls ET, Treves A, Foster D, Perez-Vicente C. 1997. Simulation studies of the CA3 hippocampal subfield modelled as an attractor neural network. *Neural Networks* 10:1559–1569.
- Rolls ET, Robertson RG, Georges-François P. 1997. Spatial view cells in the primate hippocampus. *Eur J Neurosci* 9:1789–1794.
- Rolls ET, Treves A, Robertson RG, Georges-François P, Panzeri S. 1998. Information about spatial view in an ensemble of primate hippocampal cells. *J Neurophysiol* 79:1797–1813.
- Rolls ET, Tovee MJ, Panzeri S. 1999. The neurophysiology of backward visual masking: information analysis. *J Cogn Neurosci* 11:335–346.
- Samsonovich A, McNaughton B. 1997. Path integration and cognitive mapping in a continuous attractor neural network model. *J Neurosci* 17:5900–5920.
- Schultz S, Rolls ET. 1999. Analysis of information transmission in the Schaffer collaterals. *Hippocampus* 9:582–598.
- Stringer SM, Rolls ET, Araujo I, Trappenberg T. 2000. Updating continuous attractor networks.
- Suzuki WA, Miller EK, Desimone R. 1997. Object and place memory in the macaque entorhinal cortex. *J Neurophysiol* 78:1062–1081.
- Treves A. 1990. Graded-response neurons and information encodings in autoassociative memories. *Phys Rev [A]* 42:2418–2430.
- Treves A. 1993. Mean-field analysis of neuronal spike dynamics. *Network* 4:259–284.
- Treves A. 1995. Quantitative estimate of the information relayed by the Schaffer collaterals. *J Comput Neurosci* 2:259–272.
- Treves A, Rolls ET. 1991. What determines the capacity of autoassociative memories in the brain? *Network* 2:371–397.
- Treves A, Rolls ET. 1992. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* 2:189–199.
- Treves A, Rolls ET. 1994. A computational analysis of the role of the hippocampus in memory. *Hippocampus* 4:374–391.
- Treves A, Rolls ET, Simmen M. 1997. Time for retrieval in recurrent associative memories. *Physica [D]* 107:392–400.
- Wallis G, Rolls ET. 1997. Invariant face and object recognition in the visual system. *Prog Neurobiol* 51:167–194.
- Wilson FAW, O'Sclaidhe SP, Goldman-Rakic PS. 1993. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* 260:1955–1958.