A Theory of Hippocampal Function in Memory

Edmund T. Rolls

Department of Experimental Psychology, University of Oxford, Oxford, England

ABSTRACT: First, what is computed by the hippocampus is considered. Based on the effects of damage to the hippocampus and neuronal activity recorded in the primate hippocampus, it is suggested that it is involved in associating together information usually originating from different cortical regions, for example, about objects and their place in a spatial environment. The rapid formation of such context-dependent memories is prototypical of memories of particular events or episodes. Second, a computational theory of how it performs this function, based on neuroanatomical and neurophysiological information about the different neuronal systems contained within the hippocampus, is described. Key hypotheses are that the CA3 pyramidal cells operate as a single autoassociation network to store new episodic information as it arrives via a number of specialized preprocessing stages from many different association areas of the cerebral cortex, and that the dentate granule cell/mossy fiber system is important particularly during learning to help to produce a new pattern of firing in the CA3 cells for each episode. The computational analysis shows how many memories could be stored in the hippocampus, and how quickly the CA3 autoassociation system would operate during recall. The analysis is then extended to show how the CA3 system could be used to recall the whole of an episodic memory when only a fragment of it is presented. It is shown how this retrieval within the hippocampus could lead to recall of neuronal activity in association areas of the cerebral neocortex similar to that present during the original episode, via modified synapses in backprojection pathways from the hippocampus to the cerebral neocortex. The recalled information in the cerebral neocortex could then by used by the neocortex in the formation of long-term memories and/or in the selection of appropriate actions.
© 1997 Wiley-Liss, Inc.

KEY WORDS: autoassociation, recall, episodic memory, cortical backprojections, retrograde amnesia

INTRODUCTION

The overall aim of this theory of hippocampal function in memory is to show quantitatively how the hippocampus could operate in memory. Specific aims are as follows:

1. To account for data on the function of the hippocampus in humans and monkeys as well as rats. Because memory deficits of, for example, where an object has been seen (an object-place association) are very evident in monkeys and humans following damage to the hippocampus, the theory addresses the role of the hippocampus in memory. The tasks impaired by damage to the hippocampus in monkeys and humans include object-place memory tasks, e.g., where on a video monitor an object was shown, or where on a tray different objects were located (Smith and Milner, 1981; Gaffan and Saunders, 1985; Parkinson et al., 1988; Gaffan, 1994; Angeli et al., 1993). This type of memory can be formed rapidly, in one trial, and because typically it involves associating together many aspects of the situation, such as the spatial context and the objects present, it is the type of memory used to store individual events in a snapshot-like way. An example of the type of event or episodic memory meant here is the memory of seeing a particular person at a particular place on a particular occasion. The theory thus addresses how the memory for many different events or episodes can be stored in the brain and later retrieved from a part of the episode. (This is called here episodic memory, and is not used here to refer to a whole linked sequence of different events. Linked sequences of events can be introduced at an extension of the type of event memory described here implemented by autoassociation, but are not dealt with here; cf. Levy, 1996.) The theory is thus consistent with spatial view cells in monkeys, which respond when the monkey looks toward a particular part of space (Rolls and O'Mara, 1995; Rolls, 1996a,b). These could provide the spatial context or part of an episodic memory, and indeed some monkey hippocampal neurons do respond to a combination of an object and where it was seen (Rolls et al., 1989), in allocentric coordinates (Feigenbaum and Rolls, 1991).

The theory is not a theory of spatial computation performed by the hippocampus, because although the hippocampal damage does impair spatial and episodic memory, there is little evidence that spatial computation per se is impaired after hippocampal damage in primates including humans. Instead, damage to neocortical areas such as the right parahippocampal cortex can produce deficits such as topographical disorientation (Habib and Sirigu, 1987), and damage to the right parieto-temporo-occipital region can produce topographical agnosia (see

Accepted for publication July 15, 1996.
Address correspondence and reprint requests to Prof. E.T. Rolls, Dept. of Experimental Psychology, Univ. of Oxford, South Parks Road, Oxford OX1 3UD, England.

Grüsser and Landis, 1991), both of which can be taken as requiring spatial computation. This theory thus holds that spatial computation is performed by the neocortex, and that the hippocampus may be involved in spatial computation only insofar as new information may be required to be stored or recalled in order to perform a particular spatial task. It is also noted at the outset that basing a theory of hippocampal function on its possible role as a spatial computer using place cells of the type found in rats may not generalize well to primates, in which cells which respond to spatial views "out there" are a prominent feature in the actively locomoting monkey, but cells which respond to the place where the animal is are not (Rolls and O'Mara, 1995; Rolls, 1996a,b).

2. To identify the computational functions of different parts of the hippocampal formation (CA3, CA1, dentate granule cells), and its related structures such as the parahippocampal and perirhinal cortex and the entorhinal cortex.

3. To identify the functions of the different synaptic inputs to different types of neuron in the hippocampal formation (e.g., the mossy fiber, recurrent collateral, and perforant path inputs to the CA3 cells) and to show quantitatively how the numbers of each type of synapse, and neuron, are related to the computational functions being performed.

4. To produce a quantitative theory of why there are as many backprojection pathways from the hippocampus to the neocortex (and also between adjacent neocortical areas) as there are forward projections, as well as a theory of what the backprojections perform.

5. To link studies at the cellular level of the hippocampus, including studies of the biophysics of individual cells, the rules of synaptic modification, and the information represented by hippocampal neurons in different animals when the hippocampus is functioning normally in the behaving animal, through a theory of how large numbers of such neurons could operate in a series of linked networks, to understand *how* the systems-level functions of the hippocampus are performed.

6. To show the possible relevance for computation in memory systems of different aspects of synaptic modification, including synaptic strengthening and weakening (whether or not long-term potentiation [LTP] and long-term depression [LTD] provide a good model of these processes).

7. To show how the storage of information in the hippocampus may complement that stored in other brain regions, and to consider how information may be recalled from the hippocampus, through, for example, the backprojection pathways, to the neocortex, for use in neocortical information processing and storage.

8. To show how the hippocampus could operate relatively quickly (e.g., in 50–100 ms), even though it may perform recurrent processing, and its principal neurons, pyramidal cells, fire relatively slowly, at, e.g., 0–15 spikes/s.

9. To advance at the generic level (i.e., independently of whether in the hippocampus or not) the formal understanding of how networks of neurons in the brain could operate.

The historical development of the theory described begins with Marr's work in 1971, in which having heard L. Weiskrantz lecture in 1967 on the functions of the hippocampus in memory and amnesia, he developed a mathematical model of the hippocampus with binary neurons and binary synapses which utilized heavily the properties of the binomial distribution. The article he produced was pioneering in bringing to the fore the importance of the quantitative neuroanatomy of the hippocampal system, although he himself did not attempt to specify particular functions for different cell groups within the hippocampus. A re-examination of the theory has shown that the recurrent connections between his "P3" neurons are not as crucial in his model as he thought (Willshaw and Buckingham, 1990). Gardner-Medwin (1976) showed how progressive recall could operate in a network of binary neurons with binary synapses and suggested that this might be relevant to hippocampal function. Rolls (1987) produced a theory of the hippocampus in which the CA3 neurons operated as an autoassociation memory to store episodic memories including object and place memories and the dentate granule cells operated as a preprocessing stage for this by performing pattern separation, so that the mossy fibers could act to set up different representations for each memory to be stored in the CA3 cells. He suggested that the CA1 cells operate as a recoder for the information recalled from the CA3 cells to a partial memory cue, so that the recalled information would be represented more efficiently to enable recall, via the backprojection synapses, of activity in the neocortical areas similar to that which had been present during the original episode. This theory was developed further (Rolls, 1989a–c, 1990a,b), including further details about how the backprojections could operate (Rolls, 1989a,b), and how the dentate granule cells could operate as a competitive network (Rolls, 1989c). Quantitative aspects of the theory were then developed with A. Treves, who brought the expertise of theoretical physics, applied previously mainly to understand the properties of fully connected attractor networks with binary neurons (Hopfield, 1982; Amit, 1989), to bear on the much more diluted connectivity of the recurrent collaterals found in real biological networks (e.g., 2% between CA3 pyramidal cells in the rat), in networks of neurons with graded (continuously variable) firing rates, graded synaptic strengths, and sparse representations in which only a small proportion of the neurons is active at any one time, as is found in the hippocampus (Treves, 1990; Treves and Rolls, 1991). These developments in understanding quantitatively the operation of more biologically relevant recurrent networks with modifiable synapses were applied quantitatively to the CA3 region (Treves and Rolls, 1991), and to the issue of why there are separate mossy fiber and perforant path inputs to the CA3 cells of the hippocampus (Treves and Rolls, 1992). The whole model of the hippocampus was described in more detail, and a quantitative treatment of the theory of recall by backprojection pathways in the brain was provided by Treves and Rolls (1994). The speed of operation of the CA3 system has been addressed in a number of new developments (Treves, 1993; Simmen et al., 1996b) (see below). Rolls (1995) produced a simulation of the operation of all of the hippocampus from the entorhinal cortex through the dentate, CA3, and CA1 cells back to the hippocampus, which established the quantitative feasibility of the whole theory, and raised a number of important issues considered below, including the role of topography within parts of the hippocampal internal connectivity. The simulation also emphasized some of the advantages, for a system which must store many different memories, of a binary representation, in which for any one memory the neurons were either firing or not, as opposed to having continuously graded firing rates. The simulation (Rolls, 1995) also showed how recall, if not perfect at the stage of the CA3 cells, was improved by associative synapses at subsequent stages, including the connections of the CA3 cells to the CA1 cells, and the connections of the CA1 cells to the entorhinal cortex cells. At the same time, neurophysiological investigations of the activity of neurons in the hippocampus of the actively locomoting monkey are revealing quantitatively the value of the sparseness of the firing rates of primate hippocampal neurons, which is an important parameter to the model, and is revealing the properties of "spatial view" neurons, which provide a representation that would be very appropriate for an object-place memory or an episodic memory in which spatial context is an important component (Rolls, 1996a,b).

Predictions made by the theory, and tests of the theory, are described at the end of this paper.

THE THEORY

Systems-Level Function of the Hippocampus

Any theory of the hippocampus must state at the systems level what is computed by the hippocampus. Some of the relevant evidence comes from the effects of damage to the hippocampus, the responses of neurons in the hippocampus during behavior, and the systems-level connections of the hippocampus.

Evidence from the effects of damage to the hippocampus

Damage to the hippocampus or to some of its connections such as the fornix in monkeys produces deficits in learning about the places of objects and about the places where responses should be made. 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 (Smith and Milner, 1981; Gaffan and Saunders, 1985; Parkinson et al., Murray and Mishkin, 1988; Gaffan, 1994; Angeli et al., 1993; for further review, see Rolls, 1996a,b).

Damage to the perirhinal cortex, which receives from high-order association cortex and has connections to the hippocampus (see Fig. 1), accounts for the deficits in "recognition" memory (i.e., for stimuli seen recently) produced by damage to this brain region (Zola-Morgan et al., 1989; 1994). Given that some topographic segregation is maintained in the afferents to the hippocampus through the perirhinal, parahippocampal, and entorhinal cortices (Amaral and Witter, 1989; Suzuki and Amaral, 1994), it may be that these areas are able to subserve memory within one of these topographically separated areas, of, for example, visual object, or spatial, or olfactory information. In contrast, I hypothesize that the final convergence afforded by the hippocampus into one network in CA3 (see Fig. 1) may enable the hippocampus proper to implement an event or episodic memory typically involving arbitrary associations between any of the inputs to the hippocampus, e.g., spatial, visual object, olfactory, and auditory (see below).

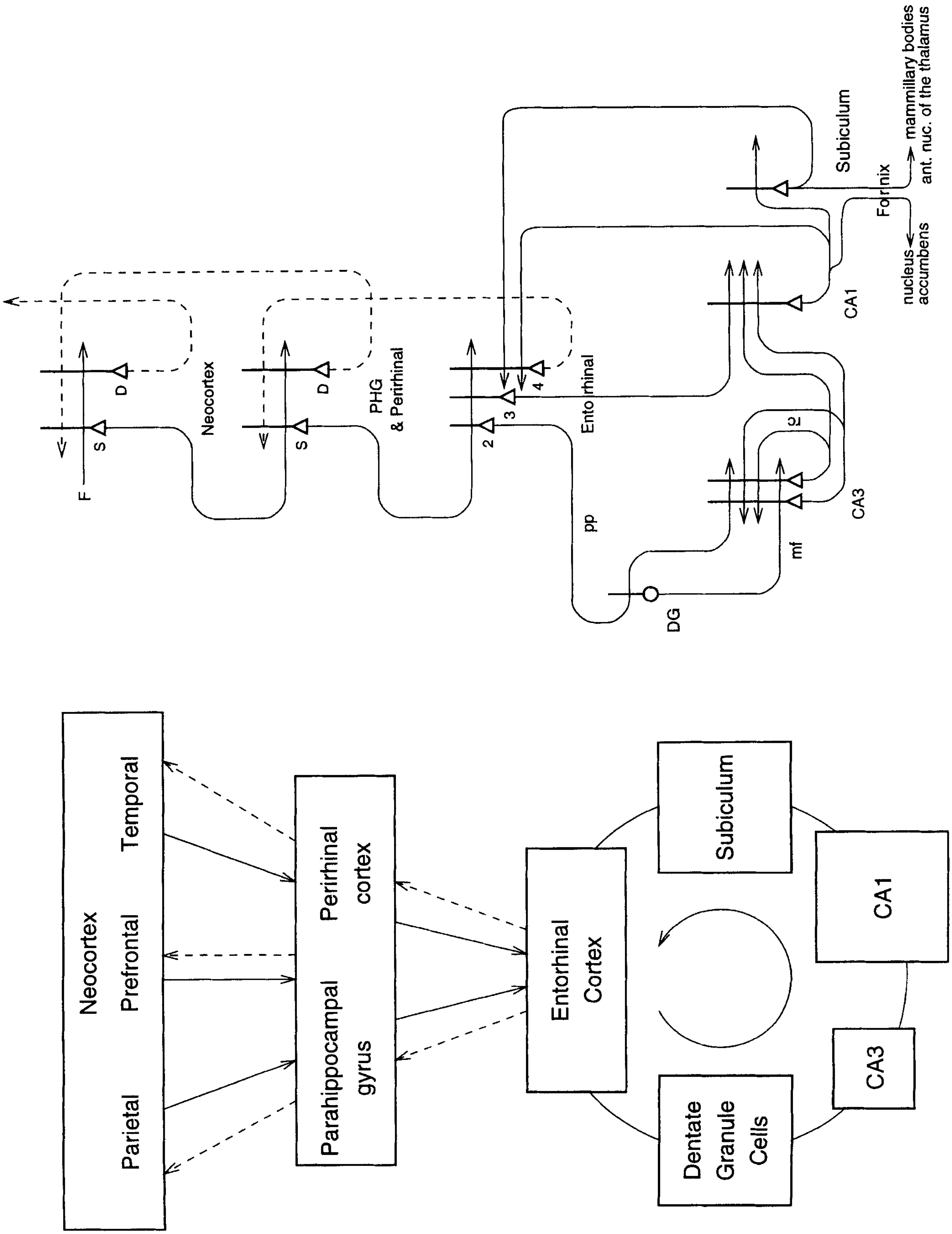
One way of relating the impairment of spatial processing to other aspects of hippocampal function is to note that this spatial processing involves a snapshot type of memory, in which one whole scene must be remembered. This memory may then be a special case of episodic memory, which involves an arbitrary association of a particular set of events which describe a past episode. Further, the nonspatial tasks impaired by damage to the hippocampal system may be impaired because they are tasks in which a memory of a particular episode or context rather than of a general rule is involved (Gaffan et al., 1984). Further, the deficit in paired associate learning in humans may be especially evident when this involves arbitrary associations between words, for example, window-lake.

I suggest that the reason why the hippocampus is used for the spatial and nonspatial types of memory described above, and the reason that makes these two types of memory so analogous, is that the hippocampus contains one stage, the CA3 stage, which acts as an autoassociation memory. It is suggested that an autoassociation memory implemented by the CA3 neurons equally enables whole (spatial) scenes or episodic memories to be formed, with a snapshot quality which depends on the arbitrary associations which can be made and the short temporal window which characterizes the synaptic modifiability in this system (see below and Rolls, 1987, 1989a,b, 1990a,b). The hypothesis is that the autoassociation memory enables arbitrary sets of concurrent activity, involving for example the spatial context where an episode occurred, the people present during the episode, and what was seen during the episode, to be associated together and stored as one event. Later recall of that episode from the hippocampus in response to a partial cue can then lead to reinstatement of the activity in the neocortex that was originally present during the episode. The theory described here shows how the episodic memory could be stored in the hippocampus and later retrieved to the neocortex.

Although there is insufficient space to review the rat literature on the effects of hippocampal damage, much of it is consistent with the evidence from primates in that spatial reference memory is impaired by hippocampal damage; this involves arbitrary associations of environmental cues to define places and, in addition, what may be present in those places (see, e.g., Jarrard, 1993). The theory described here is thus intended to be as relevant to rodents as primates, the main difference being that the spatial representation used in the hippocampus of rats would be about the place where the rat is, whereas the spatial representation in primates would be of space "out there."

The necessity to recall information from the hippocampus

The information about episodic events recalled from the hippocampus could be used to help form semantic memories (Rolls,



mammillary bodies
ant. nuc. of the thalamus
Subiculum
Fornix
nucleus
accumbens
CA1
CA3
DG
mf
pp
Ento rhinal
4
3
2
PHG
& Perirhinal
D
S
Neocortex
D
S
F
rc
Neocortex
Parietal
Prefrontal
Temporal
Parahippocampal
gyrus
Perirhinal
cortex
Entorhinal
Cortex
Dentate
Granule
Cells
CA3
CA1
Subiculum

1989a,b, 1990b; Treves and Rolls, 1994). For example, remembering many particular journeys could help to build a geographic cognitive map in the neocortex. The hippocampus and neocortex would thus be complementary memory systems, with the hippocampus being used for rapid, "on the fly," unstructured storage of information involving activity potentially arriving from many areas of the neocortex, whereas the neocortex would gradually bind and adjust on the basis of much accumulating information the semantic representation (Rolls, 1989a–c; Treves and Rolls, 1994; McClelland et al., 1995).

This raises the issue of the possible gradient of retrograde amnesia following hippocampal damage. The issue of whether memories stored some time before hippocampal damage are less impaired than more recent memories, and whether the time course is minutes, hours, days, weeks, or years, is still a debated issue (Squire, 1992; Gaffan, 1993). (In humans, there is evidence for a gradient of retrograde amnesia; in rats and monkeys, hippocampal damage in many studies appears to impair previously learned hippocampal-type memories, suggesting that in these animals, at least with the rather limited numbers of different memories that need to be stored in the tasks used, the information remains in the hippocampus for long periods.) If there is a gradient of retrograde amnesia related to hippocampal damage, then this suggests that information may be retrieved from the hippocampus if it is needed, allowing the possibility of incorporating the retrieved information into neocortical memory stores. If on the other hand there is no gradient of retrograde amnesia related to hippocampal damage, but old as well as recent memories of the hippocampal type are stored in the hippocampus and are lost if it is damaged, then again this implies the necessity of a mechanism to retrieve information stored in the hippocampus, and to use this retrieved information to affect neural circuits elsewhere (for if this were not the case, information stored in the hippocampus could never be used for anything). The current perspective is thus that whichever view of the gradient of retrograde amnesia is correct, information stored in the hippocampus will need to be retrieved and affect other parts of the brain in order to be used. The present theory shows how information could be retrieved within the hippocampus, and how this retrieved information could enable the activity in neocortical areas that was present during the original storage of the episodic event to be reinstated, thus implementing recall. The backprojections from the hippocampus to the neocortex are one of the two major outputs of the hippocampus (see Fig. 1). The backprojections are most likely to be involved in what is described by humans as recall, and in enabling information about an episode captured on the fly to be incorporated into long-term, possibly semantic, neocortical stores with a rich associative structure (cf McClelland et al., 1995). As a result of such neocortical recall, action may be initiated. The other major set of outputs from the hippocampus projects via the fimbria/fornix system to the anterior nucleus of the thalamus (both directly and via the mammillary bodies), which in turn projects to the cingulate cortex. This may provide an output for more action-directed use of information stored in the hippocampus, for example, in the initiation of conditional spatial responses in a visual conditional spatial response task (Rupniak and Gaffan, 1987; Miyashita et al., 1989). In such a task, a rapid mapping must be learned between a visual stimulus and a spatial response, and a new mapping must be learned each day. The hippocampus is involved in this rapid visual-to-spatial response mapping (Rupniak and Gaffan, 1987), and the way in which hippocampal circuitry may be appropriate for this is that the CA3 region enables signals originating from very different parts of the cerebral cortex to be associated rapidly together (see below).

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 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. The thick lines above the cell bodies represent the dendrites.

Systems-level neurophysiology of 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 hippocampus has been reviewed recently by Rolls (1996a,b), and only a brief summary can be provided here.

The primate hippocampus contains spatial cells that respond when the monkey looks at a certain part of space, for example, 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 the hippocampal neurons have such spatial view fields, and about 2.4% combine information about the 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). 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. It can be shown with the monkey stationary that these cells respond when the monkey's eye position results in him looking at a particular part of space. Some of these spatial view cells respond when the view details are obscured by curtains or darkness, when the monkey's eyes look toward the spatial view field of the neuron. In this situation, it seems appropriate to suggest that the neurons are responding to the recalled spatial view, with the recall triggered by the partial information provided about the spatial view by eye and head position information, which these experiments clearly show influence primate hippocampal neurons (experiments of E.T. Rolls, R.G. Robertson, and P. Georges-François, in preparation; see Rolls, 1996a,b, for review). 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 very 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, 1996a,b). It is therefore suggested 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. Simply looking at the place and seeing an object, or person 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 that 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. Part of the present theory is that the CA3 neurons provide an autoassociative network appropriate for rapidly (in one trial) learning such associations. The present 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.

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., 1989, 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 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, for example, return to the starting place) and to determine whether one was facing toward a particular location in space; that is, they would provide an important input for "spatial view" neurons. When spatial view neurons respond in the dark as a monkey is rotated toward a spatial view (experiments of E.T. Rolls, R.G. Robertson and P. Georges-François, in preparation; see Rolls, 1996a,b), their activity may be triggered by whole-body motion cell input into the CA3 autoassociative network.

Systems-level anatomy

The hippocampus receives, via the adjacent parahippocampal gyrus and entorhinal cortex, inputs from virtually all association areas in the neocortex, including those in the parietal, temporal, and frontal lobes (Van Hoesen, 1982; Squire et al., 1989; see Fig. 1). Therefore the hippocampus has available highly elaborated multimodal information, which has already been processed extensively along different, and partially interconnected, sensory pathways. Given that some topographic segregation is maintained in the afferents to the hippocampus through the perirhinal and parahippocampal cortices (Amaral and Witter, 1989), it may be that these areas are able to subserve memory within one of these topograpically separated areas, of, for example, visual object, *or* spatial, *or* olfactory information. In contrast, the final convergence afforded by the hippocampus into one network in CA3 (see Fig. 1) may be especially appropriate for an episodic memory typically involving arbitrary associations between any of the inputs to the hippocampus, e.g., spatial, visual object, olfactory, and auditory (see below). Additional inputs come from the amygdala and, via a separate pathway, from the septal cholinergic and other regulatory systems. An extensively divergent system of output projections enables the hippocampus to feed back into most of the cortical areas from which it receives inputs (see Fig. 1).

The operation of hippocampal circuitry as a memory system

Given the systems-level hypothesis about what the hippocampus performs, and the neurophysiological evidence about what is represented in the primate hippocampus, the next step is to consider how using its internal connectivity and synaptic modifiability the hippocampus could store and retrieve many memories, and how retrieval within the hippocampus could lead to retrieval of the activity in the neocortex that was present during the original learning of the episode. To develop understanding of how this is achieved, we have developed a computational theory of the operation of the hippocampus (see Rolls, 1987, 1989a–c, 1990a,b; Treves and Rolls, 1991, 1992, 1994). This theory, and new developments in it, are outlined next.

Hippocampal circuitry (see Fig. 1 and Storm-Mathiesen et al., 1990; Amaral and Witter, 1989; Amaral, 1993)

Projections from the entorhinal cortex reach the granule cells (of which there are 10^6 in the rat) in the dentate gyrus (DG) via the perforant path (pp). The granule cells project to CA3 cells via the mossy fibers (mf), which provide a *sparse* but possibly powerful connection to the $3 \cdot 10^5$ CA3 pyramidal cells in the rat. Each CA3 cell receives approximately 50 mossy fiber inputs, so that the sparseness of this connectivity is thus 0.005%. By contrast, there are many more, possibly weaker, direct perforant path inputs into each CA3 cell in the rat, of the order of $4 \cdot 10^3$. The largest number of synapses (about $1.2 \cdot 10^4$ in the rat) on the dendrites of CA3 pyramidal cells is, however, provided by the (recurrent) axon collaterals of CA3 cells themselves (rc). It is remarkable that the recurrent collaterals are distributed to other CA3 cells throughout the hippocampus (Ishizuka et al., 1990; Amaral and Witter, 1989; Amaral et al., 1990), so that effectively the CA3 system provides a single network, with a connectivity of approximately 2% between the different CA3 neurons given that the connections are bilateral.

CA3 as an autoassociation memory

Many of the synapses in the hippocampus show associative modification as shown by long-term potentiation, and this synaptic modification appears to be involved in learning (see Morris, 1989). On the basis of the evidence summarized above, Rolls (1987, 1989a–c, 1990a,b, 1991) has suggested that the CA3 stage acts as an autoassociation memory which enables episodic memories to be formed and stored for an intermediate term in the CA3 network, and that subsequently the extensive recurrent collateral connectivity allows for the retrieval of a whole representation to be initiated by the activation of some small part of the same representation (the cue). The crucial synaptic modification for this is in the recurrent collateral synapses. (A description of the operation of autoassociative networks is provided by Hertz et al., 1991; and by Rolls and Treves, 1997.) The hypothesis is that because the CA3 operates effectively as a single network, it can allow arbitrary associations between inputs originating from very different parts of the cerebral cortex to be formed. These might involve associations between information originating in the temporal visual cortex about the presence of an object and information originating in the parietal cortex about where it is. We have therefore performed quantitative analyses of the storage and retrieval processes in the CA3 network (Treves and Rolls, 1991, 1992). We have extended previous formal models of autoassociative memory (see Amit, 1989) by analyzing a network with graded response units, so as to represent more realistically the continuously variable rates at which neurons fire, and with incomplete connectivity (Treves, 1990; Treves and Rolls, 1991). We have found that in general the maximum number p_{max} of firing patterns that can be (individually) retrieved is proportional to the number C^{RC} of (associatively) modifiable 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

$$a = (\Sigma_{i=1,n} r_i/n)^2 / \Sigma_{i=1,n}(r_i^2/n) \tag{1}$$

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,

$$p_{\max} \cong \frac{C^{RC}}{a \ln(1/a)} \mathrm{k} \tag{2}$$

where k is a factor that depends weakly on the detailed structure of the rate distribution, on the connectivity pattern, etc., but is roughly in 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. This analysis emphasizes the utility of having a sparse representation in the hippocampus, for this enables many different memories to be stored. Third, in order for most associative networks to store information efficiently, heterosynaptic LTD (as well as LTP) is required (see Rolls and Treves, 1990; Treves and Rolls, 1991; Rolls, 1996c). Simulations that are fully consistent with the analytic theory are provided by Simmen et al. (1996a) and Rolls et al. (1997).

We have also indicated how to estimate I, the total amount of information (in bits per synapse) that can be retrieved from the network. I is defined with respect to the information i_p (in bits per cell) contained in each stored firing pattern, by subtracting the amount i_l lost in retrieval and multiplying by p/C^{RC}:

$$I = \frac{p}{C^{RC}} (i_p - i_l) \tag{3}$$

The maximal value I_{max} of this quantity was found (Treves and Rolls, 1991) to be in several interesting cases around 0.2–0.3 bits per synapse, with only a mild dependency on parameters such as the sparseness of coding a.

We may then estimate (Treves and Rolls, 1992) how much information has to be stored in each pattern for the network to efficiently exploit its information retrieval capacity I_{max}. The estimate is expressed as a requirement on i_p:

$$i_p > a \ln(1/a) \tag{4}$$

As the information content of each stored pattern i_p depends on the storage process, we see how the retrieval capacity analysis, coupled with the notion that the system is organized so as to be an efficient memory device in a quantitative sense, leads to a constraint on the storage process.

We note that although there is some spatial gradient in the CA3 recurrent connections, so that the connectivity is not fully uniform (Ishizuka et al., 1990), nevertheless the network will still have the properties of a single interconnected autoassociation network, allowing associations between arbitrary neurons to be formed, given the presence of many long-range connections which overlap from different CA3 cells.

A number of points deserve comment. First, if it is stated that a certain number of memories is the upper limit of what could be stored in a given network, then the question is sometimes asked, What constitutes a memory? The answer is precise. Any one memory is represented by the firing rates of the population of neurons that are stored by the associative synaptic modification and can be correctly recalled later. The firing rates in the primate hippocampus might be constant for a period of, for example, 1 s in which the monkey was looking at an object in one position in space; synaptic modification would occur in this time period (cf. the time course of LTP, which is sufficiently rapid for this), and the memory of the event would have been stored. The quantitative analysis shows how many such random patterns of rates of the neuronal population can be stored and later recalled correctly. If the rates were constant for 5 s while the rat was at one place or the monkey was looking at an object at one position in space, then the memory would be for the pattern of firing of the neurons in this 5-s period. (The pattern of firing of the population refers to the rate at which each neuron in the population of CA3 neurons is firing.) Second, the question sometimes arises of whether the CA3 neurons operate as an attractor network. An attractor network is one in which a stable pattern of firing is maintained once it has been started. Autoassociation networks trained with modified Hebb rules can store the number of different memories, each one expressed as a stable attractor, indicated in Equation 2. However, the hippocampal CA3 cells do not necessarily have to operate as a stable attractor: instead, it would be sufficient for the present theory if they can retrieve stored information in response to a partial cue initiating retrieval. The partial cue would remain on during recall, so that the attractor network would be operating in the clamped condition (see Rolls and Treves, 1997). The completion of the partial pattern would then provide more information than entered the hippocampus, and the extra information retrieved would help the next stage to operate. Demonstrations of this by simulations that are fully consistent with the analytic theory are provided by Rolls (1995); Simmen et al. 1996a; Rolls et al. (1997). Third, in order for most associative networks to store information efficiently, heterosynaptic LTD (as well as LTP) is required (see Rolls and Treves, 1990; Treves and Rolls, 1991; Rolls, 1996c). Without heterosynaptic LTD, there would otherwise always be a correlation between any set of positively firing inputs acting as the input pattern vector to a neuron. LTD effectively enables the average firing of each input axon to be subtracted from its input at any one time, reducing the average correlation between different pattern vectors to be stored to a low value (see Rolls, 1996c).

Given that the memory capacity of the hippocampal CA3 system is limited, it is necessary to have some form of forgetting in this store, or another mechanism to ensure that its capacity is not exceeded. (Exceeding the capacity can lead to a loss of much of the information retrievable from the network.) Heterosynaptic LTD could help this forgetting, by enabling new memories to overwrite old memories (see Rolls, 1996c). The limited capacity of the CA3 system does also provide one of the arguments that some transfer of information from the hippocampus to neocortical memory stores may be useful (see Treves and Rolls, 1994). Given its limited capacity, the hippocampus might be a useful store for only a limited period, which might be on the order of days, weeks, or months. This period may well depend on the acquisition rate of new episodic memories. If the animal were in a constant and limited environment, then as new information is not being added to the hippocampus, the representations in the hippocampus would remain stable and persistent. These hypotheses have clear experimental implications, both for recordings from single neurons and for the gradient of retrograde amnesia, both of which might be expected to depend on whether the environment is stable or frequently changing. They show that the conditions under which a gradient of retrograde amnesia might be demonstrable would be when large numbers of new memories are being acquired, not when only a few memories (few in the case of the hippocampus being less than a few hundred) are being learned.

The potential link to the gradient of retrograde amnesia is that the retrograde memories lost in amnesia are those not yet consolidated in longer-term storage (in the neocortex). As they are still held in the hippocampus, their number has to be less than the storage capacity of the (presumed) CA3 autoassociative memory. Therefore the time gradient of the amnesia provides not only a measure of a characteristic time for consolidation, but also an upper bound on the rate of storage of new memories in CA3. For example, if one were to take as a measure of the time gradient in the monkey, say, 5 weeks (about 50,000 min; Squire, 1992) and as a reasonable estimate of the capacity of CA3 in the monkey e.g., $p = 50{,}000$, then one would conclude that there is an upper bound on the rate of storage in CA3 of not more than one new memory per minute, on average. (This might be an average over many weeks; the fastest rate might be closer to 1 per s; see Treves and Rolls, 1994.)

The current theory shows how single events could be stored in and later recalled from the CA3 network. It does not explicitly deal with how a series of events could be linked together to represent a time-linked episodic sequence. To the extent that the memory of episodes involves links between successive events, this could be implemented in a network such as CA3 by introducing time delays into the (e.g., recurrent collateral) circuitry, and indeed such an approach has been discussed by Levy (1996). Difficult problems that must be considered by such an approach include how many different such sequentially linked episodic memories could be stored in the hippocampus, and whether the recall can be made to operate slowly rather than cycling through all the linked events rapidly.

The dynamics of the recurrent network

The analysis described above of the capacity of a recurrent network such as the CA3 considered steady-state conditions of the firing rates of the neurons. The question arises of how quickly the recurrent network would settle into its 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, 1992b, 1994a).

A partial answer to this question can be inferred from a recent theoretical development based on the analysis of the collective dynamical properties of realistically modeled 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 modeled 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 (Simmen et al., 1996b; Treves et al., 1997). With respect to the CA3 network, the indication is thus that retrieval would be rapid, indeed fast enough for it to be biologically plausible.

Mossy fiber inputs to the CA3 cells

We hypothesize that the mossy fiber inputs force efficient information storage by virtue of their strong and sparse influence on the CA3 cell firing rates (Rolls, 1987, 1989a,b; Treves and Rolls, 1992). (The strong effects likely to be mediated by the mossy fibers were also emphasized by McNaughton and Morris, 1989; and McNaughton and Nadel, 1990.) We hypothesize that the mossy fiber input appears to be particularly appropriate in several ways. First of all, the fact that mossy fiber synapses are large and located very close to the soma makes them relatively powerful in activating the postsynaptic cell (this should not be taken to imply that a CA3 cell can be fired by a single mossy fiber excitatory postsynaptic potential [EPSP]). Second, the firing activity of granule cells appears to be very sparse (Jung and McNaughton, 1992), and this, together with the small number of connections on each CA3 cell, produces a sparse signal, which can then be transformed into an even sparser firing activity in CA3 by a threshold effect.[1] Third, nonassociative plasticity of mossy fibers (see Brown et al., 1989, 1990) might have a useful effect in enhancing the signal-to-noise ratio, in that a consistently firing mossy fiber would produce nonlinearly amplified currents in the postsynaptic cell, which would not happen with an occasionally firing fiber (Treves and Rolls, 1992). This plasticity, and also learning in the dentate, would also have the effect that similar fragments of each episode (e.g., the same environmental location) recurring on subsequent occasions would be more likely to activate the same population of CA3 cells, which would have potential advantages in terms of economy of use of the CA3 cells in different memories, and in making some link between different episodic memories with a common feature, such as the same location in space. Fourth, with only a few, and powerful, active mossy fiber inputs to each CA3 cell, setting a given sparseness of the representation provided by CA3 cells would be simplified, for the EPSPs produced by the mossy fibers would be Poisson distributed with large membrane potential differences for each active mossy fiber. Setting the average firing rate of the dentate granule cells would effectively set the sparseness of the CA3 representation, without great precision being required in the threshold setting of the CA3 cells (Rolls and Perez-Vicente, in preparation). Part of what is achieved by the mossy fiber input may be setting the sparseness of the CA3 cells correctly, which, as shown above, is very important in an autoassociative memory store.

The argument based on information suggests, then, that an input system with the characteristics of the mossy fibers is essential during learning, in that it may act as a sort of (unsupervised) teacher that effectively strongly influences which CA3 neurons fire based on the pattern of granule cell activity. This establishes an information-rich neuronal representation of the episode in the CA3 network (see Treves and Rolls, 1992). The quantitative analysis shows that the perforant path input would not produce a pattern of firing in CA3 that contains sufficient information for learning (Treves and Rolls, 1992).

On the basis of these points, we predict that the mossy fibers may be necessary for new learning in the hippocampus but may not be necessary for recall of existing memories from the hippocampus. Experimental evidence consistent with this prediction about the role of the mossy fibers in learning has been described in mice without mossy fiber LTP associated with a lack of the mGluR1 receptor (Conquet et al., 1994).

If acetyl choline does turn down the efficacy of the recurrent collateral synapses between CA3 neurons (Hasselmo et al., 1995), then cholinergic activation also might help to allow external inputs rather than the internal recurrent collateral inputs to dominate the firing of the CA3 neurons during learning, as the current theory proposes. If cholinergic activation at the same time facilitated LTP in the recurrent collaterals (as it appears to in the neocortex), then cholinergic activation could have a useful double role in facilitating new learning at times of behavioral activation, when presumably it may be particularly relevant to allocate some of the limited memory capacity to new memories.

Perforant path inputs to CA3 cells

By calculating the amount of information that would end up being carried by a CA3 firing pattern produced solely by the perforant path input and by the effect of the recurrent connections, we have been able to show (Treves and Rolls, 1992) that an input of the perforant path type, alone, is unable to direct efficient information storage. Such an input is too weak, it turns out, to drive the firing of the cells, as the "dynamics" of the network is dominated by the randomizing effect of the recurrent collaterals. This is the manifestation, in the CA3 network, of a general prob-

[1] For example, if only one granule cell in 100 were active in the dentate gyrus, and each CA3 cell received a connection from 50 randomly placed granule cells, then the number of active mossy fiber inputs received by CA3 cells would follow a Poisson distribution of average 50/100 = 1/2, i.e., 60% of the cells would not receive any active input, 30% would receive only one, 7.5% two, little more than 1% would receive three, and so on. (It is easy to show from the properties of the Poisson distribution and our definition of sparseness that the sparseness of the mossy fiber signal as seen by a CA3 cell would be $x/(1 + x)$, with $x = C^{MF}a_{DG}$, assuming equal strengths for all mossy fiber synapses.) If three mossy fiber inputs were required to fire a CA3 cell and these were the only inputs available, we see that the activity in CA3 would be roughly as sparse, in the example, as in the dentate gyrus.

lem affecting storage (i.e., learning) in *all* autoassociative memories. The problem arises when the system is considered to be activated by a set of input axons making synaptic connections that have to compete with the recurrent connections, rather than having the firing rates of the neurons artificially clamped into a prescribed pattern.

An autoassociative memory network needs afferent inputs also in the other mode of operation, i.e., when it retrieves a previously stored pattern of activity. We have shown (Treves and Rolls, 1992) that the requirements on the organization of the afferents are in this case very different, implying the necessity of a second, separate input system, which we have identified with the perforant path to CA3. In brief, the argument is based on the notion that the cue available to initiate retrieval might be rather small, i.e., the distribution of activity on the afferent axons might carry a small correlation, $q \ll 1$, with the activity distribution present during learning. In order not to lose this small correlation altogether, but rather transform it into an input current in the CA3 cells that carries a sizable signal—which can then initiate the retrieval of the full pattern by the recurrent collaterals—one needs a large number of associatively modifiable synapses. This is expressed by the formulas that give the specific signal S produced by sets of associatively modifiable synapses, or by nonassociatively modifiable synapses: If C^{AFF} is the number of afferents per cell

$$S_{\mathrm{ASS}} \sim \frac{\sqrt{C^{\mathrm{AFF}}}}{\sqrt{p}}\,q \qquad S_{\mathrm{NONASS}} \sim \frac{1}{\sqrt{C^{\mathrm{AFF}}}}\,q. \tag{5}$$

Associatively modifiable synapses are therefore needed, and are needed in a number of C^{AFF} of the same order as the number of concurrently stored patterns p, so that small cues can be effective; whereas nonassociatively modifiable synapses—or even more so, nonmodifiable ones—produce very small signals, which decrease in size the larger the number of synapses. In contrast with the storage process, the average strength of these synapses does not now play a crucial role. This suggests that the perforant path system is the one involved in relaying the cues that initiate retrieval.

Before leaving the CA3 cells, it is suggested that separate scaling of the three major classes of excitatory input to the CA3 cells (recurrent collateral, mossy fiber, and perforant path, see Fig. 1) could be independently scaled, by virtue of the different classes of inhibitory interneuron which receive their own set of inputs, and end on different parts of the dendrite of the CA3 cells (Gulyas et al., 1993; cf. for CA1 Buhl et al., 1994). This possibility is made simpler by having these major classes of input terminate on different segments of the dendrites. Each of these inputs, and the negative feedback produced through inhibitory interneurons when the CA3 cells fire, should for optimal functioning be separately regulated (Rolls, 1995), and the anatomical arrangement of the different types of inhibitory interneuron might be appropriate for achieving this.

Dentate granule cells

The theory is developed elsewhere that the dentate granule cell stage of hippocampal processing which precedes the CA3 stage acts in four ways to produce during learning the sparse yet efficient (i.e., nonredundant) representation in CA3 neurons which is required for the autoassociation to perform well (Rolls, 1989a–c, 1994b; see also Treves and Rolls, 1992).

The first way is that the perforant path–dentate granule cell system with its Hebb-like modifiability is suggested to act as a competitive learning network to remove redundancy from the inputs producing a more orthogonal, sparse, and categorized set of outputs (Rolls, 1987, 1989a–c, 1990a,b). The nonlinearity in the NMDA receptors may help the operation of such a competitive net, for it ensures that only the most active neurons left after the competitive feedback inhibition have synapses that become modified and thus learn to respond to that input (Rolls, 1989c). We note that if the synaptic modification produced in the dentate granule cells lasts for a period of more than the duration of learning the episodic memory, then it could reflect the formation of codes for regularly occurring combinations of active inputs that might need to participate in different episodic memories. Because of the nonlinearity in the NMDA receptors, the nonlinearity of the competitive interactions between the neurons (produced by feedback inhibition and nonlinearity in the activation function of the neurons) need not be so great (Rolls, 1989c). Because of the feedback inhibition, the competitive process may result in a relatively constant number of strongly active dentate neurons relatively independently of the number of active perforant path inputs to the dentate cells. The operation of the dentate granule cell system as a competitive network may also be facilitated by a Hebb rule of the form:

$$\delta w_{\mathrm{ij}} = k \cdot r_{\mathrm{i}}\,(r'_{\mathrm{j}} - \mathrm{w}_{\mathrm{ij}}) \tag{6}$$

were k is a constant, r_{i} is the activation of the dendrite (the postsynaptic term), r'_{j} is the presynaptic firing rate, w_{ij} is the synaptic weight, and r'_{j} and w_{ij} are in appropriate units (see Rolls, 1989c). Incorporation of a rule such as this which implies heterosynaptic LTD as well as LTD (see Levy and Desmond, 1985; Levy et al., 1990) makes the sum of the synaptic weights on each neuron remain roughly constant during learning (cf. Oja, 1982; see Rolls, 1989c).

The second way is also a result of the competitive learning hypothesized to be implemented by the dentate granule cells (Rolls, 1987; 1989a–c, 1990a,b, 1994b). It is proposed that this allows overlapping (or very similar) inputs to the hippocampus to be separated, in the following way (see also Rolls, 1994a,b). Consider three patterns B, W, and BW where BW is a linear combination of B and W. (To make the example very concrete, we could consider binary patterns where B = 10, W = 01, and BW = 11.) Then the memory system is required to associate B with reward, W with reward, but BW with punishment. This is one of the configural learning tasks of Sutherland and Rudy (1991), and for them is what characterizes the memory functions performed by the hippocampus. Without the hippocampus, rats might have more difficulty in solving such problems. However, it is a property of competitive neuronal networks that they can separate such overlapping patterns, as has been shown elsewhere (Rolls, 1989c; Rolls and Treves, 1997; normalization of synaptic weight vectors is required for this property). It is thus an important part of hippocampal neuronal network architecture that there is a competitive network that precedes the CA3 autoassociation system. Without the dentate gyrus, if a conventional autoassociation network were presented with the mixture BW having learned B and W separately, then the autoassociation network would produce a mixed output state, and would therefore be incapable of storing separate memories for B, W, and BW. It is suggested therefore that competition in the dentate gyrus is one of the powerful computational features of the hippocampus, which could enable it to help solve what have been called configural types of learning task (Sutherland and Rudy, 1991). (It is a separate and perhaps not fully resolved issue of the extent to which the ability to solve configural learning tasks is a crucial and distinguishing role of the hippocampus in memory. It is suggested that such all or none characterizations may be less useful than understanding that computationally the separation of overlapping patterns before storage in memory is a function to which the hippocampus may be particularly able to contribute because of the effect just described. It is not inconsistent with this if configural learning can take place without the hippocampus [see Rudy and Sutherland, 1995]; one might just expect it to be better with the hippocampus, particularly when a large number of such overlapping memories must be stored and retrieved.)

The third way arises because of the very low contact probability in the mossy fiber-CA3 connections, and has been explained above in the section Mossy Fibers Inputs to the CA3 Cells and by Treves and Rolls (1992).

A fourth way is that, as suggested and explained above in the section just mentioned, the dentate granule cell-mossy fiber input to the CA3 cells may be powerful, and its use particularly during learning would be efficient in forcing a new pattern of firing onto the CA3 cells during learning.

CA1 cells

The amount of information about each episode retrievable from CA3 has to be balanced off against the number of episodes that can be held concurrently in storage. The balance is regulated by the sparseness of the coding. Whatever the amount of information per episode in CA3, one may hypothesize that the organization of the structures that follow CA3 (i.e., CA1, the various subicular fields, and the return projections to neocortex) should be optimized so as to preserve and use this information content in its entirety. This would prevent further loss of information, after the massive but necessary reduction in information content that has taken place along the sensory pathways and before the autoassociation stage in CA3. We have proposed (Treves and Rolls, 1994; Treves, 1995) that the need to preserve the full information content present in the output of an autoassociative memory requires an intermediate recoding stage (CA1) with special characteristics. In fact, a calculation of the information present in the CA1 firing pattern, elicited by a pattern of activity retrieved from CA3, shows that a considerable fraction of the information is lost if the synapses are nonmodifiable, and that this loss can be prevented only if the CA3 to CA1 synapses are associatively modifiable. Their modifiability should match the plasticity of the CA3 recurrent collaterals. The additional information that can be retrieved beyond that retrieved by CA3 because the CA3 to CA1 synapses are associatively modifiable is strongly demonstrated by the hippocampal simulation described by Rolls (1995).

An additional factor is that if the total amount of information carried by CA3 cells is redistributed over a larger number of CA1 cells, less information needs to be loaded onto each CA1 cell, rendering the code more robust to information loss in the next stages. For example, if each CA3 cell had to code for 2 bits of information, e.g., by firing at one of four equiprobable activity levels, then each CA1 cell (if there were twice as many as there are CA3 cells) could code for just 1 bit, e.g., by firing at one of only two equiprobable levels. Thus the same information content could be maintained in the overall representation while reducing the sensitivity to noise in the firing level of each cell. In fact, there are more CA1 cells than CA3 cells in rats (2.5×10^5). There are even more CA1 cells (4.6×10^6) in humans (and the ratio of CA1 to CA3 cells is greater). The CA1 cells may thus provide the first part of the expansion for the return projections to the enormous numbers of neocortical cells in primates, after the bottleneck of the single network in CA3, the number of neurons in which may be limited because it has to operate as a single network.

Another argument on the operation of the CA1 cells is also considered to be related to the CA3 autoassociation effect. In this, several arbitrary patterns of firing occur together on the CA3 neurons, and become associated together to form an episodic or "whole-scene" memory. It is essential for this operation that several different sparse representations are present conjunctively in order to form the association. Moreover, when completion operates in the CA3 autoassociation system, all the neurons firing in the original conjunction can be brought into activity by only a part of the original set of conjunctive events. For these reasons, a memory in the CA3 cells consists of several different simultaneously active ensembles of activity. To be explicit, the parts A, B, C, D, and E of a particular episode would each be represented, roughly speaking, by its own population of CA3 cells, and these five populations would be linked together by autoassociation. It is suggested that the CA1 cells, which receive these groups of simultaneously active ensembles, can detect the conjunctions of firing of the different ensembles which represent the episodic memory, and allocate by competitive learning neurons to represent at least larger parts of each episodic memory (Rolls, 1987, 1989a–c, 1990a,b). In relation to the simple example above, some CA1 neurons might code for ABC, and others for BDE, rather than having to maintain independent representations in CA1 of A, B, C, D, and E. This implies a more efficient representation, in the sense that when eventually after many further stages, neocortical neuronal activity is recalled (as discussed below), each neocortical cell need not be accessed by all the axons carrying each component A, B, C, D, and E, but instead by fewer axons carrying larger fragments, such as ABC, and BDE. Concerning the details of operation of the CA1 system, we note that although competitive learning may capture part of how it is able to recode, the competition is probably not global, but instead would operate relatively locally within the domain of the connections of inhibitory neurons. This simple example is intended to show how the coding may become less componential and more conjunctive in CA1 than in CA3, but should not be taken to imply that the representation produced becomes more sparse.

Another feature of the CA1 network is its double set of afferents, with each of its cells receiving most synapses from the Schaeffer collaterals coming from CA3, but also a proportion (about 1/6; Amaral et al., 1990) from direct perforant path projections from entorhinal cortex. Such projections appear to originate mainly in layer 3 of entorhinal cortex (Witter et al., 1989), from a population of cells only partially overlapping with that (mainly in layer 2) giving rise to the perforant path projections to DG and CA3. This suggests that it is useful to include in CA1 not only what is possible to recall from CA3, but also the detailed information present in the retrieval cue itself (see Treves and Rolls, 1994).

Backprojections to the neocortex—a hypothesis

The need for information to be retrieved from the hippocampus to affect other brain areas was noted in the Introduction. The way in which this could be implemented via backprojections to the neocortex is now considered.

It is suggested that the modifiable connections from the CA3 neurons to the CA1 neurons allow the whole episode in CA3 to be produced in CA1. This may be assisted as described above by the direct perforant path input to CA1. This might allow details of the input key for the recall process, as well as the possibly less information-rich memory of the whole episode recalled from the CA3 network, to contribute to the firing of CA1 neurons. 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, 1989a–c). 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 from which it is thought that cortical areas such as 39 and 40 relate to language developed), and also areas of unimodal association cortex (e.g., inferior temporal visual cortex). The backprojections, by recalling previous episodic events, could provide information useful to the neocortex in the building of new representations in the multimodal and unimodal association cortical areas (Rolls, 1989a–c, 1990a,b), or in organizing actions.

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 we suggest that the backprojection synapses are set up to have the appropriate strengths for recall is as follows (see also Rolls, 1989a,b). During the setting up of a new episodic memory, there would be strong feedforward activity progressing toward 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 neocortex, that is 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 (ITC) neurons become activated by visual stimuli with latencies of 90–110 ms and may continue firing for several hundred milliseconds (Rolls, 1992b); 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, 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, 1989a,b].)

The synapses receiving the backprojections would have to be Hebb-modifiable, as suggested by Rolls (1989a,b). This would solve the de-addressing problem, that is 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, 1989a,b) 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, 1989a,b), 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 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 follows from the operation of the system described above is that there would be no point, indeed it would be disadvantageous, if the synaptic modifications lasted for longer than the memory remained in the hippocampal buffer store. What would be optimal would be to arrange for the associative modification of the backprojecting synapses to remain for as long as the memory persists in the hippocampus. This suggests that a similar mechanism for the associative modification within the hippocampus and for that of at least one stage of the backprojecting synapses would be appropriate. 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 (i.e., rapid learning with 1–2 s, and slow decay over perhaps weeks) 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 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, 1989a,b) 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.

We have noted elsewhere that backprojections, which included cortico-cortical backprojections, and backprojections originating from structures such as the hippocampus and amygdala, may have a number of different functions (Rolls, 1989a–c, 1990a,b, 1992a). The particular function with which we have been concerned here is how memories stored in the hippocampus might be recalled in regions of the cerebral neocortex.

Backprojections to the neocortex—quantitative aspects

How many backprojecting fibers does one need to synapse on 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 lines 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 and have been indeed modified at the time of first experiencing each episode, 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 backward toward 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 equation (2) above

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

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 p is equal to the number of memories held in the hippocampal memory, it is limited by the retrieval capacity of the CA3 network, p_{max}. Putting together the formula for the latter with that shown here, one concludes that, roughly, the requirement implies that the number of afferents of (indirect) hippocampal origin to a given neocortical stage (C^{HBP}) must be $C^{HBP} = C^{RC} a_{nc} / a_{\mathrm{CA3}}$, where C^{RC} is the number of recurrent collaterals to any given cell in CA3, the average sparseness of a representation is a_{nc}, and a_{CA3} is the sparseness of memory representations there in CA3.

The above requirement is very strong: even if representations were to remain as sparse as they are in CA3, which is unlikely, to avoid degrading the signal, C^{HBP} should be as large as C^{RC}, i.e. 12,000 in the rat. Moreover, other sources of noise not considered in the present calculation would add to the severity of the constraint and partially compensate for the relaxation in the constraint that would result from requiring that only a fraction of the p episodes would involve any given cortical area. If then C^{HBP} has to be of the same order as C^{RC}, one is led to a very definite conclusion: A mechanism of the type envisaged here 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., C^{HBP} times the ratio between the number of neocortical and that 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).

Although backprojections between any two adjacent areas in the cerebral cortex are approximately as numerous as forward projections, and much of the distal parts of the dendrites of cortical pyramidal cells are devoted to backprojections, the actual number of such connections onto each pyramidal cell may be on average only in the order of thousands. Further, not all 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 (allowing, for example, 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 the 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. (Note that this completion would be only over the local information present within a cortical area about, e.g., visual input *or* spatial input; it provides a local "clean-up" mechanism, and could not replace the global autoassociation performed effectively over the activity of very many cortical areas which the CA3 could perform by virtue of its widespread recurrent collateral connectivity.) There are two alternative possibilities about how this would operate. First, if the recurrent collaterals showed slow and long-lasting synaptic modification, then they would be useful in completing the whole of long-term (e.g., semantic) memories. Second, if the neocortical recurrent collaterals showed rapid changes in synaptic modifiability with the same time course as that of hippocampal synaptic modification, then they would be useful in filling in parts of the information forming episodic memories which could be made available locally within an area of the cerebral neocortex.

Simulations of hippocampal operation

In order to test the operation of the whole system for individual parts of which an analytic theory has now been developed, Rolls (1995) simulated a scaled down version of the part of the architecture shown in Figure 1 from the entorhinal cortex to the hippocampus and back to the entorhinal cortex. The analytic approaches to the storage capacity of the CA3 network, the role of the mossy fibers and of the perforant path, the functions of CA1, and the operation of the backprojections in recall were all shown to be computationally plausible in the computer simulations. In the simulation, during recall, partial keys are presented to the entorhinal cortex, completion is produced by the CA3 autoassociation network, and recall is produced in the entorhinal cortex of the original learned vector. The network, which has 1,000 neurons at each stage, can recall large numbers, which approach the calculated storage capacity, of different sparse random vectors. One of the points highlighted by the simulation is that the network operated much better if the CA3 cells operated in binary mode (either firing or not), rather than having continuously graded firing rates (Rolls, 1995). The reason for this is that given that the total amount of information that can be stored in a recurrent network such as the CA3 network is approximately constant independently of how graded the firing rates are in each pattern (Treves, 1990), then if much information is used to store the graded firing rates in the firing of CA3 cells, fewer patterns can be stored. The implication of this is that in order to store many memories in the hippocampus, and to be able to recall them at later stages of the system, for example, the entorhinal cortex and beyond, it may be advantageous to utilize relatively binary firing rates in the CA3 part at least of the hippocampus. This finding has been confirmed and clarified by simulation of the CA3 autoassociative system alone, and it has been suggested that the advantage of operation with binary firing rates may be related to the low firing rates characteristic of hippocampal neurons (Rolls et al., 1997). Another aspect of the theory emphasized by the results of the simulation was the importance of having effectively a single network provided in the hippocampus by the CA3 recurrent collateral network, for only if this operated as a single network (given the constraint of some topography present at earlier stages), could the whole of a memory be completed from any of its parts.

DISCUSSION

Quantitative Aspects of the Model

The model described here is quantitative and is supported by both formal analyses and quantitative simulations. Many of the points made, such as on the number of memories that can be stored in autoassociative networks, the utility of sparse representations, and the dynamics of the operation of networks with recurrent connections, are quite general and will apply to networks in a number of different brain areas. With respect to the hippocampus, the theory specifies the maximum number of memories that could be stored in it, and this has implications for how it could be used biologically. It indicates that if this number is approached, it will be useful to have a mechanism for recalling information from the hippocampus for incorporation into memories elsewhere. With respect to recall, the theory provides a quantitative account for why there are as many backprojections as forward projections in the cerebral cortex. Overall, the theory provides an explanation for *how* this part of the brain could work, and even if this theory needs to be revised, it is suggested that a fully quantitative theory along the lines proposed which is based on the evidence available from a wide range of techniques will be essential before we can say that we understand *how* a part of the brain operates.

Comparison With Other Theories of Hippocampal Function

Hypotheses have been described about how a number of different parts of hippocampal and related circuitry might operate. Although these hypotheses are consistent with a theory of how the hippocampus operates, some of these hypotheses could be incorporated into other views or theories. In order to highlight the differences between alternative theories, and in order to lead to constructive analyses that can test them, the theory described above is compared with other theories of hippocampal function in the following section. Although the differences between the theories are highlighted in the following section, the overall view described here is close in different respects to those of a number of other investigators (Marr, 1971; Brown and Zador, 1990; McNaughton and Nadel, 1990; Eichenbaum et al., 1992; Gaffan, 1992; Squire, 1992), and of course priority is not claimed on all the propositions put forward here.

Some theories postulate that the hippocampus performs spatial computation. The theory of O'Keefe and Nadel (1978), that the hippocampus implements a cognitive map, placed great emphasis on spatial function. It supposed that the hippocampus at least holds information about allocentric space in a form which enables rats to find their way in an environment even when novel trajectories are necessary, that is, it permits an animal to "go from one place to another independent of particular inputs (cues) or outputs (responses), and to link together conceptually parts of the environment which have never been experienced at the same time." O'Keefe (1990; see Burgess et al., 1994) has extended this analysis and produced a computational theory of the hippocampus as a cognitive map, in which the hippocampus performs geometric spatial computations. Key aspects of the theory are that the hippocampus stores the centroid and slope of the distribution of landmarks in an environment and stores the relationships between the centroid and the individual landmarks. The hippocampus then receives as inputs information about where the rat currently is, and where the rat's target location is, and computes geometrically the body turns and movements necessary to reach the target location. In this sense, the hippocampus is taken to be a spatial computer which produces an output which is very different from its inputs. This is in contrast to the present theory, in which the hippocampus is a memory device, which is able to recall what was stored in it, using as input a partial cue. The theory of O'Keefe postulates that the hippocampus actually performs a spatial computation.

McNaughton et al. (1991) have also proposed that the hippocampus is involved in spatial computation. They propose a "compass" solution to the problem of spatial navigation along novel trajectories in known environments, postulating that distances and bearings (i.e., vector quantities) from landmarks are stored, and that computation of a new trajectory involves vector subtraction by the hippocampus. They postulate that a linear associative mapping is performed, using as inputs a "cross-feature" (combination) representation of (head) angular velocity and (its time integral) head direction, to produce as output the future value of the integral (head direction) after some specified time interval. The system can be reset by learned associations between local views of the environment and head direction, so that when later a local view is seen, it can lead to an output from the network which is a (corrected) head direction. They suggest that some of the key signals in the computational system can be identified with the firing of hippocampal cells (e.g., local view cells) and subicular cells (head direction cells). It should be noted that this theory requires a (linear) associative mapping with an output (head direction) different in form from the inputs (head angular velocity over a time period, or local view). This is pattern association (with the conditioned stimulus local view, and the unconditioned stimulus head direction), not autoassociation, and it has been postulated that this pattern association can be performed by the hippocampus (cf. McNaughton and Morris, 1989). This theory is again in contrast to the present theory, in which the hippocampus operates as a memory to store events that occur at the same time, and can recall the whole memory *from any part* of what was stored. (A pattern associator uses a conditioned stimulus to map an input to a pattern of firing in an output set of neurons which is like that produced in the output neurons by the unconditioned stimulus. A description of pattern associations and autoassociators in a neurobiological context is provided by Rolls, 1996c; and Rolls and Treves, 1997.) The present theory is fully consistent with the presence of "spatial view" cells and whole-body motion cells in the primate hippocampus (Rolls and O'Mara, 1993) (or place or local view cells in the rat hippocampus, and head direction cells in the presubiculum), for it is often important to store and later recall where one has been (views of the environment, body turns made, etc), and indeed such (episodic) memories are required for navigation by "dead reckoning" in small environments.

The present theory thus holds that the hippocampus is used for the formation of episodic memories using autoassociation. This function is often necessary for successful spatial computation, but is not itself spatial computation. Instead, we believe that spatial computation is more likely to be performed in the parietal cortex (utilizing information recalled from the hippocampus if necessary). Consistent with this view, hippocampal damage impairs the ability to learn new environments but not to perform spatial computations such as finding one's way to a place in a familiar environment, whereas damage to the parietal cortex and parahippocampal cortex can lead to problems such as topographical and other spatial agnosias in humans (see Kolb and Whishaw, 1990; Grusser and Landis, 1991). This is consistent with spatial computations normally being performed in the neocortex. (In monkeys, there is evidence for a role of the parietal cortex in allocentric spatial computation. For example, monkeys with parietal cortex lesions are impaired at performing a landmark task, in which the object to be chosen is signified by the proximity to it of a "landmark" (another object) (Ungerleider and Mishkin, 1982).)

Another theory was sketched by Marr (1971). He had the general systems-level view, to which we subscribe, that the hippocampal system operates as an intermediate-term memory. His theory, however, did not identify functions for different parts of the hippocampal circuitry (dentate, CA3, CA1, subiculum, etc.), but instead lumped them together. He discussed the possible functions of associatively modifiable recurrent collateral connections, but in the quantities he assumed in his model, synaptic modification on the forward synaptic connections into the system, rather than in the recurrent collaterals, was quantitatively significant in the storage effects analyzed (Willshaw and Buckingham, 1990). The technical approach Marr took to the analysis was based on probabilities computed in the tail of Poisson distributions, and as noted by Treves and Rolls (1994), the conclusions reached with this approach are strongly affected by details of the assumptions made (e.g., how many extra active inputs to a cell are needed to make it fire?). In contrast, we have adopted a more powerful analytic approach, based on formal models derived from theoretical physics of the operation of attractor neuronal networks, which we have extended in the direction of biological plausibility by incorporating linear-threshold rather than binary neurons in sparse networks with nonsymmetric synaptic weights. The methods we use also introduce information theory to the assessment of how different input systems operate in such attractor networks (see above and Treves and Rolls, 1994). A second contrast of the approach that we have adopted is that we have, given that there is now much more information available on hippocampal function (from, for example, microanatomy and neurophysiology), been able to address the possible specific functions of several stages of processing in the hippocampal system. A third contrast is that Marr suggested that memories might be unloaded from the hippocampus to the cerebral neocortex during sleep (a suggestion taken up by Wilson and McNaughton, 1994). On the other hand, in the approach taken here, it is suggested that if episodic information stored in the hippocampus is recalled to the neocortex to help build long-term semantic memories, a process which may often require small modifications to synaptic weights in the light of new episodic information, a serial process guided by thinking about how the new episodic information is related to existing semantic or long-term episodic information is more likely to be required (cf. McClelland et al., 1995). A fourth contrast is that although Marr (1971) promised a theory of how information could be recalled from the hippocampus to the neocortex, he did not as far as I know ever produce such a theory. Rolls (1989a,b) and Treves and Rolls (1994) have outlined a theory of recall and have provided what may be some strong quantitative constraints on the system in the brain which achieves this. This system is identified with the multistage backprojection system from the hippocampus to the cerebral neocortex, and between adjacent neocortical areas.

Another theory is that the hippocampus is involved in recognition memory (Mishkin, 1978, 1982). It is now believed that recognition memory as tested in a visual delayed-match-to-sample task is dependent on the perirhinal cortex, and rather less on hippocampal circuitry proper (Zola-Morgan et al., 1989; Gaffan and Murray, 1992). Our approach to this is that we note that the hippocampal CA3 recurrent collateral system is most likely to be heavily involved in memory processing when new associations between arbitrary events which may be represented in different regions of the cerebral cortex must be linked together to form an episodic memory. Often, given the large inputs to the hippocampus form the parietal cortex, one of these events will contain spatial information. We suppose that given the circuitry of the hippocampus, it is especially well suited for such tasks, although some mixing of inputs may occur before the hippocampus. It is therefore predicted that when arbitrary associations must be rapidly learned between such different events to form new episodic memories, the hippocampal circuitry is likely to become increasingly important, but we are not surprised if some memory formation of this type can occur without the hippocampus proper. The position implies that hippocampal damage will reflect quantitatively rather than just qualitatively the ability to form new (especially multimodal, with one modality space) episodic memories. It is also noted that Mishkin's theory was a theory of what the hippocampus does, whereas the present theory is a theory of what the hippocampus does and especially of how it does it.

A theory closely related to the present theory of how the hippocampus operates has been developed by McClelland et al. (1995). It is very similar to the theory we have developed (Rolls, 1987, 1989a–c; Treves and Rolls, 1992, 1994) at the systems level, except that it takes a stronger position on the gradient of retrograde amnesia, emphasizes that recall from the hippocampus of episodic information is used to help build semantic representations in the neocortex, and holds that the last set of synapses that are modified rapidly during the learning of each episode are those between the CA3 and CA1 pyramidal cells (see Fig. 1). In the formulation by McClelland et al (1995), the entorhinal cortex connections via the perforant path onto the CA1 cells are nonmodifiable (in the short term) and allow a representation of neocortical long-term memories to activate the CA1 cells. The new information learned in an episode by the CA3 system is then linked to existing long-term memories by the CA3-to-CA1 rapidly modifiable synapses. All the connections from the CA1 back via the subiculum, entorhinal cortex, parahippocampal cortex, etc. to the association neocortex are held to be unmodifiable in the short term, during the formation of an episodic memory. The formal argument that leads us to suggest that the backprojecting synapses *are* associatively modifiable during the learning of an episodic memory is similar to that which we have used to show that for efficient recall, the synapses which initiate recall in the CA3 system (identified above with perforant path projection to CA3) must be associatively modifiable if recall is to operate efficiently (see Treves and Rolls, 1992). The present theory holds that it is possible that for several stages back into neocortical processing, the backprojecting synapses should be associatively modifiable, with a similar time course to the time it takes to learn a new episodic memory. It may well be that at earlier stages of cortical processing, for example, from V4 to V2, the backprojections are relatively more fixed, being formed during early developmental plasticity or during the formation of new long-term semantic memory structures. Having such relatively fixed synaptic strengths in these earlier cortical backprojection systems could ensure that whatever is recalled in higher cortical areas, such as objects, will in turn recall relatively fixed and stable representations of parts of objects or features. Given that the functions of backprojections may include many top-down processing operations, including attention and priming, it may be useful to ensure that there is consistency in how higher cortical areas affect activity in earlier "front-end" or preprocessing cortical areas.

If a model of the hippocampal/neocortical memory system could store only a small number of patterns, it would not be a good model of the real hippocampal/neocortical memory system in the brain. Indeed, this appears to be a major limitation of another model presented recently by Alvarez and Squire (1994). The model specifies that the hippocampus helps the operation of a neocortical multimodel memory system in which all memories are stored by associative recurrent collaterals between the neocortical neurons. Although the idea worked in the model with 20 neurons and two patterns to be learned (Alvarez and Squire, 1994), the whole idea is computationally not feasible, because the number of memories that can be stored in a single autoassociative network of the type described is limited by the number of inputs per neuron from other neurons, not by the number of neurons in the network (Treves and Rolls, 1991, 1994). This would render the capacity of the whole neocortical multimodal (or amodal) memory store very low (in the order of the number of inputs per neuron from the other neurons, that is in the order of 5,000–10,000) (cf. O'Kane and Treves, 1992). This example makes it clear that it is important to take into account analytic and quantitative approaches when investigating memory systems. The current work is an attempt to do this.

The aim of this comparison of the present theory with other theories has been to highlight differences between the theories, to assist in the future assessment of the utility and the further development of each of the theories.

Predictions

Some of the main predictions arising from the present theory are brought together here. All can in principle be tested experimentally.

1. The recurrent collateral and perforant path synaptic systems to CA3 should display associative modifiability.
2. Blocking this modifiability should impair the formation of new (hippocampal-dependent) memories, but should not impair the retention of previously stored memories. (The impairment is likely to be most demonstrable when large numbers of episodic memories are to be stored.)
3. Selective inactivation of the mossy fiber system should impair memory formation but not memory retention.
4. The associative plasticity of the Schaffer collaterals onto CA1 cells should be similar, in strength and time course, to that of the recurrent collaterals onto CA3 cells.
5. The backprojecting system from the hippocampus must be associatively modifiable for at least one stage, and should operate as rapidly (i.e., within 1–2 s) as the associative modifiability within the hippocampus itself, and should decay (if at all) with a similar slow time course (e.g., weeks) to that of CA3 LTP. This should hold for at least one stage in the series of backprojecting stages from the CA3 region to the neocortex.
6. Neocortical cells activated solely by backprojecting inputs should have the same response characteristics as when they are activated directly by the feedforward inputs.

In addition, the quantitative analysis predicts a series of detailed quantitative relationships that will be testable once more refined experimental techniques allow more precise quantitation of e.g., lesion effects, cell response properties, and behavioral impairments. For example, the memory functions performed by the hippocampus may only become fully revealed when large numbers (thousands) of memories of particular events are required. Another example is that the information made available by the responses of each cell in CA1 should be less than that available in single neurons in the CA3 region (see Treves, 1995).

CONCLUSIONS

In this work, a number of recent neurophysiological, neuroanatomical, and theoretical investigations have been brought together to provide the outline of a theory of how the hippocampus could compute, and how the computations it performs could be used to recall recent memories. It has been shown how the recall of information within the hippocampus could lead to recall in the cerebral cortex via hippocampal backprojections, and it has been suggested how this could be useful to the cerebral cortex. A number of experimental tests of the theory have been suggested.

Acknowledgments

Different parts of the research described here were supported by the Medical Research Council, grant PG8513790; by a Human Frontier Science program grant; by an EEC BRAIN grant; by the MRC Oxford Interdisciplinary Research Centre in Brain and Behaviour; and by the Oxford McDonnell-Pew Centre in Cognitive Neuroscience. 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, Shane O'Mara, Conrado Perez-Vicente, Martin Simmen, and David Foster, and their contributions are sincerely acknowledged. Discussions with Peter Somogyi were very helpful.

REFERENCES

Alvarez P, Squire LR (1994) Memory consolidation and the medial temporal lobe: a simple network model. Proc Natl Acad Sci USA 91: 7041–7045.

Amit DJ (1989) Modelling brain function. New York: Cambridge University Press.

Amaral DG (1993) Emerging principles of intrinsic hippocampal organization. Curr Opin Neurobiol 3:225–229.

Amaral DG, Witter MP (1989) The three-dimensional organization of the hippocampal formation: a review of anatomical data. Neuroscience 31:571–591.

Amaral DG, Ishizuka N, Claiborne B (1990) Neurons, numbers and the hippocampal network. Prog Brain Res 83:1–11.

Amaral DG, Price JL, Pitkanen A, Carmichael ST (1992). Anatomical organization of the primate amygdaloid complex. In *The Amygdala,* (Aggleton JP, ed), pp 1–66. Wiley-Liss: New York.

Angeli SJ, Murray EA, Mishkin M (1993) Hippocampectomized monkeys can remember one place but not two. Neuropsychologia 31: 1021–1030.

Brown TH, Zador A (1990) The hippocampus. In: The synaptic organization of the brain (Shepherd G, ed), pp 346–388. New York: Oxford University Press.

Brown TH, Ganong AH, Kairiss EW, Keenan CL, Kelso SR (1989) Long-term potentiation in two synaptic systems of the hippocampal brain slice. In: Neural models of plasticity (Byrne JH, Berry WO, eds), pp 266–306. San Diego: Academic Press.

Brown TH, Kairiss EW, Keenan CL (1990) Hebbian synapses: biophysical mechanisms and algorithms. Annu Rev Neurosci 13:475–511.

Buhl EH, Halasy K, Somogyi P (1994) Diverse sources of hippocampal unitary inhibitory postsynaptic potentials and the number of synaptic release sites. Nature 368:823–828.

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

Cahusac PMB, Miyashita Y, Rolls ET (1989) Responses of hippocampal formation neurons in the monkey related to delayed spatial response and object-place memory tasks. Behav Brain Res 33:229–240.

Cahusac PMB, Rolls ET, Miyashita Y, Niki H (1993) Modification of the responses of hippocampus neurons in the monkey during the learning of a conditional spatial response task. Hippocampus 3:29–42.

Conquet F, Bashir ZI, Davies CH, Daniel H, Ferragutl F, Bordl F, Franz-Bacon K, Reggiani A, Matarese V, Conde F, Collingridge GL, Crepel F (1994) Motor deficit and impairment of synaptic plasticity in mice lacking mGluR1. Nature 372:237–243.

Eichenbaum H, Otto T, Cohen NJ (1992) The hippocampus—what does it do? Behav Neural Biol 57:2–36.

Feigenbaum JD, Rolls ET (1991) Allocentric and egocentric spatial information processing in the hippocampal formation of the behaving primate. Psychobiology 19:21–40.

Gaffan D (1977) Monkey's recognition memory for complex pictures and the effects of fornix transection. Q J Exp Psychol 29:505–514.

Gaffan D (1992) The role of the hippocampo-fornix-mammillary system in episodic memory. In: Neuropsychology of memory (2nd Ed) (Squire LR, Butters N, eds), pp 336–346. New York: Guilford.

Gaffan D (1993) Additive effects of forgetting and fornix transection in the temporal gradient of retrograde amnesia. Neuropsychologia 31:1055–1066.

Gaffan D (1994) Scene-specific memory for objects: a model of episodic memory impairment in monkeys with fornix transection. J Cogn Neurosci 6:305–320.

Gaffan D, Murray EA (1992) Monkeys (Macaca fascicularis) with rhinal cortex ablations succeed in object discrimination learning despite 24-hr intertrial intervals and fail at matching to sample despite double sample presentations. Behav Neurosci 106:30–38.

Gaffan D, Saunders RC, Gaffan EA, Harrison S, Shields C, Owen MJ (1984) Effects of fornix transection upon associative memory in monkeys: role of the hippocampus in learned action. Q J Exp Psychol 26B:173–221.

Gardner-Medwin AR (1976) The recall of events through the learning of associations between their parts. Proc R Soc Lond [Biol] 194:375–402.

Grusser O-J, Landis T (1991) Visual agnosias. London: MacMillan (Chapter 21).

Gulyas AI, Miles R, Hajos N, Freund TF (1993) Precision and variability in postsynaptic target selection of inhibitory cells in the hippocampal CA3 region. Eur J Neurosci 5:1729–1751.

Habib M, Sirigu A (1987) Pure topographical disorientation: a definition and anatomical basis. Cortex 23:73–85.

Hasselmo ME, Schnell E, Barkai E (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. J Neurosci 15:5249–5262.

Hertz J, Krogh A, Palmer RG (1991) Introduction to the theory of neural computation. Wokingham, UK: Addison-Wesley.

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Nat Acad Sci USA 79:2554–2558.

Ishizuka N, Weber J, Amaral DG (1990) Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. J Comp Neurol 295:580–623.

Jarrard EL (1993) On the role of the hippocampus in learning and memory in the rat. Behav Neural Biol 60:9–26.

Jung MW, McNaughton BL (1992) Spatial selectivity of unit activity in the hippocampal granular layer. Hippocampus 3:165–182.

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.

Kolb B, Whishaw IQ (1990) Fundamentals of human neuropsychology, 3rd Ed. New York: Freeman.

Levy WB, Desmond NL (1985) The rules of elemental synaptic plasticity. In: Synaptic modification, neuron selectivity, and nervous system organization (Levy WB, Anderson JA, Lehmkuhle S, eds), Ch 6, pp 105–121. Hillsdale, NJ: Erlbaum.

Levy WB (1996) Context codes and the effect of noisy learning on a simplified hippocampal CA3 model. Biol Cybern 74:159–165.

Levy WB, Colbert CM, Desmond NL (1990) Elemental adaptive processes of neurons and synapses: a statistical/computational perspective. In: Neuroscience and connectionist theory (Gluck M, Rumelhart D, eds), Ch 5, pp 187–235. Hillsdale, NJ: Erlbaum.

Marr D (1971) Simple memory: a theory for archicortex. Phila Trans R Soc Lond [Biol] 262:24–81.

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.

McNaughton BL, Nadel L (1990) Hebb-Marr networks and the neurobiological representation of action in space. In: Neuroscience and connectionist theory (Gluck MA, Rumelhart DE, eds), pp 1–64. Hillsdale, NJ: Erlbaum.

McNaughton BL, Barnes CA, Meltzer J, Sutherland RJ (1989) Hippocampal granule cells are necessary for normal spatial learning but not for spatially selective pyramidal cell discharge. Exp Brain Res 76:485–496.

McNaughton BL, Chen LL, Markus EJ (1991) "Dead reckoning," landmark learning, and the sense of direction: a neurophysiological and computational hypothesis. J Cogn Neurosci 3:190–202.

Mishkin M (1978) Memory severely impaired by combined but not separate removal of amygdala and hippocampus. Nature 273:297–298.

Mishkin M (1982) A memory system in the monkey. Philos Trans R Soc [Biol] 298:85–95.

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.

Morris RGM (1989) Does synaptic plasticity play a role in information storage in the vertebrate brain? In: Parallel distributed processing: implications for psychology and neurobiology (Morris RGM, ed), Ch 11, pp 248–285. Oxford: Oxford University Press.

Oja E (1982) A simplified neuron model as a principal component analyzer. J Math Biol 15:267–273.

O'Kane D, Treves A (1992) Why the simplest notion of neocortex as an autoassociative memory would not work. Network 3:379–384.

O'Keefe J, Nadel L (1978) The hippocampus as a cognitive map. Oxford: Clarendon Press.

O'Keefe J (1990) A computational theory of the cognitive map. Prog Brain Res 83:301–312.

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, Tamura R, Nishijo H, Nakamura K (1993) Neural mechanisms of recognition and memory in the limbic system. In: Brain mechanisms of perception and memory: from neuron to behavior (eds. Ono T, Squire LR, Raichle ME, Perrett DI, Fukuda M), ch 19, pp 330–355. New York: Oxford University Press.

Parkinson JK, Murray EA, Mishkin M (1988) A selective mnemonic role for the hippocampus in monkeys: memory for the location of objects. J Neurosci 8:4059–4167.

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

Rolls ET (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, Konishi M, eds), pp 503–540. Chichester: Wiley.

Rolls ET (1989a) Functions of neuronal networks in the hippocampus and neocortex in memory. In: Neural models of plasticity: experimental and theoretical approaches (Byrne JH, Berry WO, eds), ch 13, pp 240–265. San Diego: Academic Press.

Rolls ET (1989b) The representation and storage of information in neuronal networks in the primate cerebral cortex and hippocampus. In: The computing neuron (Durbin R, Miall C, Mitchison G, eds), ch 8, pp 125–159. Wokingham, England: Addison-Wesley.

Rolls ET (1989c) Functions of neuronal networks in the hippocampus and cerebral cortex in memory. In: Models of brain function (Cotterill RMJ, ed), pp 15–33. Cambridge: Cambridge University Press.

Rolls ET (1990a) Theoretical and neurophysiological analysis of the functions of the primate hippocampus in memory. Cold Spring Harb Symp Quant Biol 55:995–1006.

Rolls ET (1990b) Functions of the primate hippocampus in spatial processing and memory. In: Neurobiology of comparative cognition (Olton DS, Kesner RP, eds), ch 12, pp 339–362. Hillsdale, NJ: Lawrence Erlbaum.

Rolls ET (1990c) A theory of emotion, and its application to understanding the neural basis of emotion. Cognition Emotion 4:161–190.

Rolls ET (1991) Functions of the primate hippocampus in spatial and non-spatial memory. Hippocampus 1:258–261.

Rolls ET (1992a) Neurophysiology and functions of the primate amygdala. In: The amygdala (Aggleton JP, ed), ch 5, pp 143–165. New York: Wiley-Liss.

Rolls ET (1992b) Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Philos Trans R Soc Lond [Biol] 335:11–21.

Rolls ET (1994a) Brain mechanisms for invariant visual recognition and learning. Behav Proc 33:113–138.

Rolls ET (1994b) Neurophysiological and neuronal network analysis of how the primate hippocampus functions in memory. In: The memory system of the brain, (Delacour J, ed) pp 713–744. London: World Scientific.

Rolls ET (1995) A model of the operation of the hippocampus and entorhinal cortex in memory. Int J Neural Systems [Suppl] 6:51–70.

Rolls ET (1996a) The representation of space in the primate hippocampus, and episodic memory. In: Perception, memory and emotion: frontier in neuroscience (Ono T, McNaughton BL, Molotchnikoff S, Rolls ET, Nishijo H, eds), pp 375–400. Amsterdam: Elsevier.

Rolls ET (1996b) The representation of space in the primate hippocampus, and its relation to memory. In: Brain processing and memory (Sakata H, et al., eds), pp 203–227. Amsterdam: Elsevier.

Rolls ET (1996c) Roles of long term potentiation and long term depression in neuronal network operations in the brain. In: Cortical plasticity: LTP and LTD (Fazeli MS, Collingridge GL, eds), pp 223–250. Oxford: Bios.

Rolls ET, O'Mara S (1993) Neurophysiological and theoretical analysis of how the hippocampus functions in memory. In: Brain mechanisms of perception: from neuron to behavior (Ono T, Squire LR, Raichle M, Perrett D, Fukuda M, eds), ch 17, pp 276–300. New York: Oxford University Press.

Rolls ET, O'Mara SM (1995) View-responsive neurons in the primate hippocampal complex. Hippocampus 5:409–424.

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 (1997) Neural networks and brain function. 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, in press.

Rudy JW, Sutherland RJ (1995) Configural association theory and the hippocampal formation: an appraisal and reconfiguration. Hippocampus 5:375–389.

Rupniak NMJ, Gaffan D (1987) Monkey hippocampus and learning about spatially directed movements. J Neurosci 7:2331–2337.

Simmen MA, Treves A, Rolls ET (1996a) Pattern retrieval in threshold-linear associative nets. Network 7:109–122.

Simmen MW, Rolls ET, Treves A (1996b) On the dynamics of a network of spiking neurons. In: Computations and neuronal systems: Proceedings of CNS95 (Eekman FH, Bower JM, eds). Boston: Kluwer.

Smith ML, Milner B (1981) The role of the right hippocampus in the recall of spatial location. Neuropsychologia 19:781–793.

Squire LR (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys and humans. Psychol Rev 99:195–231.

Squire LR, Shimamura AP, Amaral DG (1989) Memory and the hippocampus. In: Neural models of plasticity: theoretical and empirical approaches (Byrne J, Berry WO, eds), ch 12, pp 208–239. New York: Academic Press.

Storm-Mathiesen J, Zimmer J, Ottersen OP (eds) (1990) Understanding the brain through the hippocampus. Prog Brain Res 83.

Sutherland RJ, Rudy JW (1991) Exceptions to the rule of space. Hippocampus 1:250–252.

Suzuki W, Amaral DG (1994) Topographic organisation of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. J Neurosci 14:1856–1877.

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*, in press.

Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In: Analysis of visual behaviour (Ingle D, Goodale MA, Mansfield RJW, eds), pp 549–586. Cambridge, MA: MIT Press.

Willshaw DJ, Buckingham JT (1990) An assessment of Marr's theory of the hippocampus as a temporary memory store. Philos Trans R Soc Lond [Biol] 329:205–215.

Wilson MA, McNaughton BL (1994) Reactivation of hippocampal ensemble memories during sleep. Science 265:603–604.

Witter MP, Groenewegen HJ, Lopes da Silva FH, Lohman AHM (1989)

Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region. Prog Neurobiol 33:161–254.

Van Hoesen GW (1982) The parahippocampal gyrus. New observations regarding its cortical connections in the monkey. Trends Neurosci 5:345–350.

Zola-Morgan S, Squire LR, Amaral DG, Suzuki WA (1989) Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment. J Neurosci 9:4355–4370.

Zola-Morgan S, Squire LR, Ramus SJ (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.