HIPPOCAMPUS, VOL. 4, NO. 3, PAGES 374–391, JUNE 1994

Computational Analysis of the Role of the Hippocampus in Memory

Alessandro Treves and Edmund T. Rolls

Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, England

ABSTRACT

The authors draw together the results of a series of detailed computational studies and show how they are contributing to the development of a theory of hippocampal function. A new part of the theory introduced here is a quantitative analysis of how backprojections from the hippocampus to the neocortex could lead to the recall of recent memories. The theory is then compared with other theories of hippocampal function. First, what is computed by the hippocampus is considered. The hypothesis the authors advocate, on the basis of the effects of damage to the hippocampus and neuronal activity recorded in it, is that it is involved in the formation of new memories by acting as an intermediate-term buffer store for information about episodes, particularly for spatial, but probably also for some nonspatial, information. The authors analyze how the hippocampus could perform this function, by producing a computational theory of how it operates, based on neuroanatomical and neurophysiological information about the different neuronal systems contained within the hippocampus. 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 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 a whole episodic memory when only a fragment of it is presented. It is shown how this recall could operate using modified synapses in backprojection pathways from the hippocampus to the cerebral neocortex, resulting in reinstatement of neuronal activity in association areas of the cerebral neocortex similar to that present during the original episode. The recalled information in the cerebral neocortex could then be used by the neocortex in the formation of long-term memories.

Key words: autoassociation, recall, episodic memory, cortical backprojections, retrograde amnesia

INTRODUCTION

During the last few years, we, and others, have been developing detailed hypotheses on how the hippocampus could function. Some of these investigations have led to formal analyses of neuronal networks relevant to particular aspects of how the hippocampus could perform computations. Parts of these formal analyses have been published (see e.g., Treves and Rolls, 1991, 1992). The aims of this paper are to draw together our analyses and ideas on how the hippocampus could function, and to relate and compare the resulting theory with other theories, in order to stimulate new tests of the different theories. In addition, this article presents new analyses that may help us to understand both the organization of the CA1 region, and how the hippocampus could recall recent memories in the neocortex, and thus play an important role in the consolidation of long-term memories stored in the cerebral neocortex.

Damage to the hippocampus or to some of its connections, such as the fornix in monkeys, produces deficits in learning about the places of responses and about the places of stimuli (Rolls, 1990b, 1991). For example, fornix lesions impair conditional left–right discrimination learning, in which the visual appearance of an object specifies whether a response is to be made to the left or the right (Rupniak and Gaffan, 1987). A comparable deficit is found in humans (Petrides, 1985). Macaques and humans with damage to the hippocampus or fornix are also impaired in object–place memory tasks in which not only the ob-

Address correspondence and reprint requests to Dr. E.T. Rolls, University of Oxford, Department of Experimental Psychology, South Parks Road, OX 1 3UD, Oxford, England.
Alessandro Treves is now at S.I.S.S.A. - Biofisica, via Beirut 2-4, 34103 Trieste, Italy.

jects seen, but where they were seen, must be remembered (Parkinson et al., 1988). Such object–place tasks require a whole scene or "snapshot"-like memory in which spatial relations in a scene must be remembered (Gaffan and Harrison, 1989).

Damage to the perirhinal and parahippocampal cortex part of the hippocampal system in primates produces impairments in visual recognition memory tasks including delayed match to sample (Zola-Morgan et al., 1989; see Gaffan, 1977; Squire, 1992; Rolls et al., 1993).

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 set of events that 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 rather than of a general rule is involved (Rolls, 1990a–c, 1991; Rolls and O'Mara, 1993).

THE HIPPOCAMPAL SYSTEM

Architecture

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 (Squire et al., 1989). Therefore, the hippocampus has available highly elaborated multimodal information that has already been processed extensively along different, and partially interconnected, sensory pathways. Additional inputs come from the amygdala and, via a separate pathway, from the cholinergic and other regulatory systems. An extensively divergent system of output projections enables the hippocampus to feed back into most of the areas from which it received inputs.

Information is processed within the hippocampus along a distinctly unidirectional path, consisting of three major stages, as shown in Figure 1 (see Amaral and Witter, 1989; the articles in Storm-Mathiesen et al., 1990; Amaral, 1993). Axonal projections mainly from layer 2 of the entorhinal cortex reach the granule cells in the dentate gyrus via the perforant path (PP), and also proceed to make synapses on the apical dendrites of pyramidal cells in the next stage, CA3. A different set of fibers projects from the entorhinal cortex (mainly layer 3) directly onto the third processing stage, CA1.

There are about 10^6 dentate granule cells in the rat, and more than 10 times as many in man (more detailed anatomical studies are available for the rat) (Amaral et al., 1990; West and Gundersen, 1990). They project to CA3 cells via the mossy fibers (MF), which form a relatively *sparse* but possibly powerful synaptic matrix; each fiber makes, in the rat, about 15 synapses onto the proximal dendrites of CA3 pyramidal cells. As there are some 3×10^5 CA3 pyramidal cells in the rat (Sprague-Dawley; 2.3×10^6 in man: Seress, 1988), each of them receives no more than around 50 mossy synapses. (The sparseness of this connectivity is thus 0.005%.) By contrast, there are many more—possibly weaker—direct perforant path inputs onto each CA3 cell (in the rat there are approximately 4×10^3). The largest number of synapses (about 1.2×10^4 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 (Amaral and Witter, 1989; Ishizuka et al., 1990), so that essentially the CA3 system provides a single network, with a connectivity of approximately 4% between the different CA3 neurons. The implication of this widespread recurrent collateral connectivity is that each CA3 cell can transmit information to every other CA3 cell within two or three synaptic steps. (In the rat, the CA3 fibers even connect across the midline, so that, essentially in the rat there is a single CA3 network in the hippocampus, with 2% connectivity on average between the 600,000 neurons—see Rolls, 1990b.) The CA3 system therefore is, far more than either DG or CA1, a system in which intrinsic, recurrent excitatory connections are, at least numerically, dominant with respect to excitatory afferents.

In addition, there are also intrinsic connections with a variety of numerically limited and mainly inhibitory populations of interneurons, as well as extrinsic connections with sublimbic structures; such connections are widely believed to subserve generic regulation of neural activity in CA3, as opposed to providing signals specific to the information being processed in the system.

Extrinsic axonal projections from CA3, the Schaffer collaterals, provide the major input to CA1 pyramidal cells, of which there are about 4×10^5 in the Sprague-Dawley rat. The CA1 pyramidal cells are characteristically smaller than the CA3 cells and, across different species, come in larger numbers. In terms of cell numbers, therefore, information appears to be funneled from DG through the CA3 bottleneck, and then spread out again into CA1. The output of CA1 returns via the subiculum to the entorhinal cortex, from which it is redistributed to neocortical areas.

Plasticity

Neurophysiological evidence also indicates that many of the synapses within the hippocampus are modified as a result of experience, in a way explicitly related to the types of learning for which the hippocampus is necessary, as shown in studies (e.g., Morris, 1989) in which the blocking of such modifiability with drugs results in specific learning impairments.

Studies on long-term potentiation (LTP) have shown that some synaptic systems (in DG and CA1, but possibly also PP and RC synapses in CA3) display a Hebbian, or associative, form of plasticity, whereby presynaptic activity concurrent with strong postsynaptic depolarization can result in a strengthening of the synaptic efficacy (Miles, 1988; Brown et al., 1990). Such strengthening appears to be associated with the activation of NMDA (N-Methyl-D-Aspartate) receptors (Collingridge and Singer, 1990), and it is possible that the same synapses display also (associative) long-term depression (Levy and Desmond, 1985; Levy et al., 1990). Also, MF synapses are known to display long-term activity-dependent synaptic enhancement, but this form of enhancement appears not to be associative (Brown et al., 1990).

Outline of a computational hypothesis

Rolls (1987, 1989a–c, 1990a,b, 1991) has suggested that the reason why the hippocampus is used for the spatial and non-

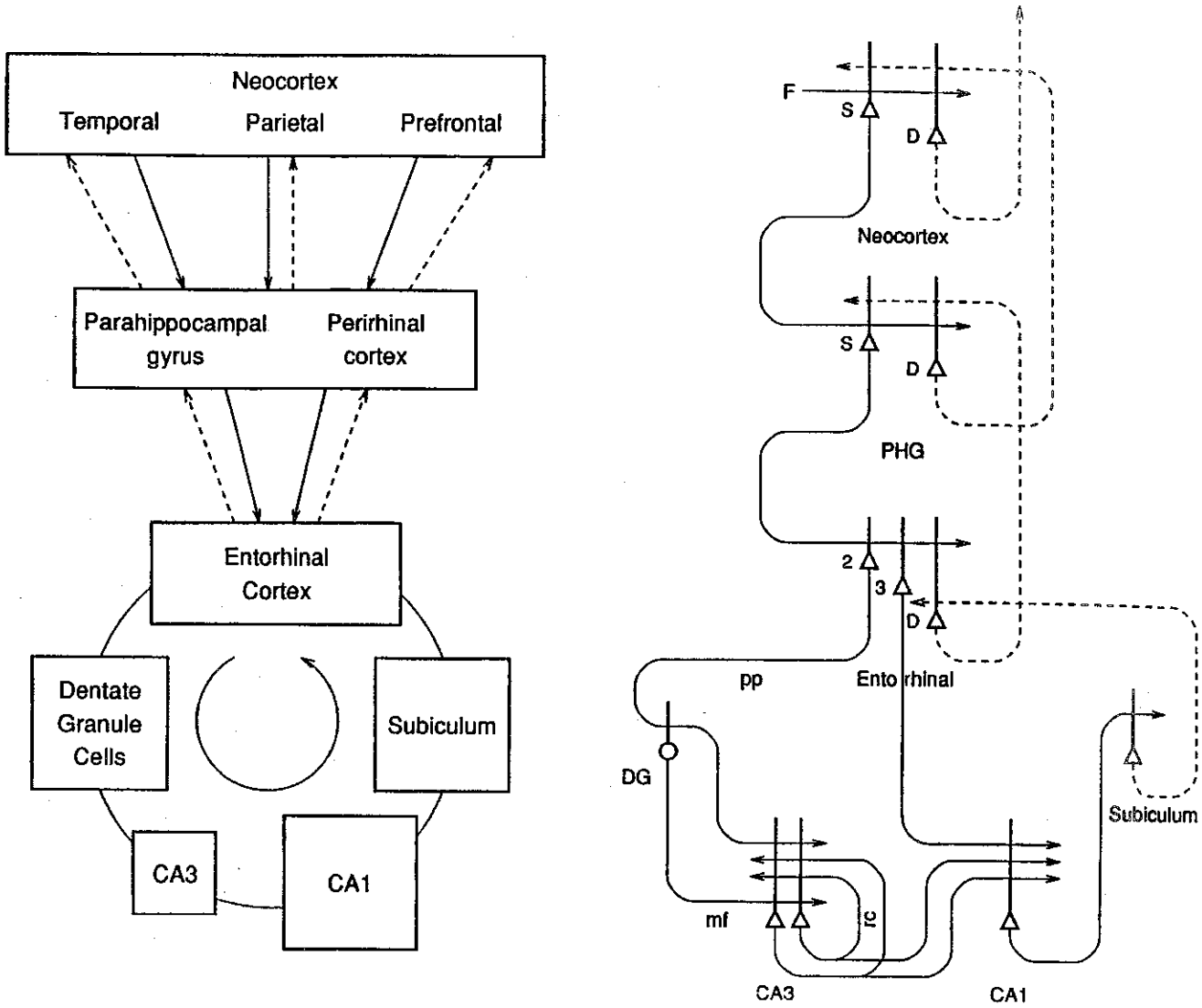

Fig. 1. Forward connections (solid lines) from areas of cerebral association neocortex via the parahippocampal and perirhinal cortex, and entorhinal cortex, to the hippocampus; and backprojections (dashed lines) via hippocampal CA1 pyramidal cells, subiculum, and parahippocampal gyrus to the neocortex. There is great convergence in the forward connections down to the single network implemented in the CA3 pyramidal cells; and great divergence again in the backprojections. **Left**: Block diagram. **Right**: More detailed representation of some of the principal excitatory neurons in the pathways. D, deep pyramidal cells; DG, dentate Granule cells; F, forward inputs to areas of the association cortex from preceding cortical areas in the hierarchy; mf, mossy fibers; PHG, parahippocampal gyrus and perirhinal cortex; pp, perforant path; rc, recurrent collateral of the CA3 hippocampal pyramidal cells; S, Superficial pyramidal cells; 2, pyramidal cells in layer 2 of the entorhinal cortex; 3, pyramidal cells in layer 3 of the entorhinal cortex. The thick lines above the cell bodies represent the dendrites.

spatial 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. An autoassociation memory implemented by the CA3 neurons would enable whole (spatial) scenes as well as episodic memories to be formed, with a snapshot quality that depends on the arbitrary associations that can be made and the short temporal window which characterizes the synaptic modifiability in this system. This hypothesis implies that any new event to be memorized is given a unitary representation as a firing pattern of CA3 pyramidal cells, that the pattern is stored into associatively modifiable synapses, and that subsequently the extensive RC 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). Cued retrieval of the CA3 representation might subserve both recall and the consolidation of more permanent memory traces, which is likely to involve slower processes, in the neocortex. Quantitative aspects of this hypothesis, leading to a detailed model of the functioning of some parts of the hippocampal circuitry, are described in the next section.

It is suggested that this ability to link information originating from different brain regions in a single autoassociation matrix in the CA3 region is a key feature of hippocampal

architecture, which is unlikely to be implemented in input regions such as the entorhinal cortex and dentate granule cells, which not only do not have the required system of recurrent collaterals over the whole population of cells, but also appear to maintain some segregation of inputs originating from different parts of the cerebral cortex (Insausti et al., 1987). Indeed, the inputs which reach the hippocampus, via the parahippocampal gyrus and the entorhinal cortex (see Fig. 1), originate from very different cortical areas, which are at the ends of the main cortical processing streams, and are not heavily interconnected. For example, these inputs come from the inferior temporal visual cortex (but not from early cortical visual areas), from the parietal cortex, from the superior temporal auditory cortex, and from the prefrontal cortex. In this anatomical sense, the hippocampus in its CA3 system provides a single net in which information from these different cortical areas could be combined to form a *snapshot*, that is, an episodic, memory (Rolls, 1987, 1989a–c, 1990a,b, 1991). Such an episodic memory might, by combining inputs from these different cortical systems contain information about what one ate at lunch the previous day, where it was, and the faces and voices of the people who were present. Information about rewards, punishments, and emotional states that can be parts of episodic memories could be incorporated into the hippocampal representation of an episodic memory via the afferents to the entorhinal cortex from structures, such as the amygdala, that are involved in processing information about rewards and emotions (Rolls, 1990c, 1992a,b). The relatively nonspecific subcortical afferents to the hippocampus, for example, the cholinergic inputs from the medial septum (which would be interrupted by the fornix section that produces memory impairments; Gaffan, 1992) may be involved in threshold setting in the hippocampus, and in particular may, by allowing arousal to affect the hippocampus, make it more likely that memories will be stored when in certain states, such as arousal (Rolls, 1989b). This would tend to economize in the use of the available storage space for episodic memories in the hippocampus, for new episodic memories would be less likely to be stored during low arousal when little may be happening in the environment, but would be stored strongly when significant and therefore arousing environmental events are occurring (see Wilson and Rolls, 1990a,b). The effects of cholinergic and other nonspecific inputs to the hippocampus may be realised at least partly because of their role in making it more likely that postsynaptic activation will exceed the threshold for activation of the N-methyl-D-aspartate (NMDA) receptors involved in long-term potentiation.

Neurophysiological evidence on how information is represented in the hippocampus

When considering the operation of networks in the brain, it is important to define what information is presented to the neuronal networks, as commented upon by Churchland and Sejnowski (1992), and also what information leaves the network. Evidence on this for the hippocampus comes not only indirectly from anatomical evidence on systems-level connectivity, but also directly from recordings of the activity of single neurons made in the hippocampus, and in the structures which provide afferents to it, when these brain regions are performing the functions for which they are required. Indeed, it is for these reasons that whole series of recordings have been made from single neurons in the hippocampus of monkeys (see Rolls, 1989a,b, 1990a,b; Rolls and O'Mara, 1993), and rats (see Kubie and Muller, 1991; Leonard and McNaughton, 1990; O'Keefe, 1990, 1991; Jung and McNaughton, 1993) performing memory and spatial tasks.

Such neurophysiological evidence shows, first, that information reaches the primate hippocampus about which visual stimuli are currently being shown, about where they are in space, and about whole body motion as signaled by vestibular inputs and by optic flow (see, e.g., Rolls and O'Mara, 1993). Second, it shows that the hippocampus is a system in which different information originating in different cortical areas can be brought together onto the same neurons. For example, in an object–place memory task in which the monkey had to remember not only which object was seen, but where on a video monitor it had appeared (Rolls et al., 1989) it was found that 9% of neurons recorded in the hippocampus and parahippocampal gyrus had spatial fields, in that they responded whenever there was a stimulus in some but not in other positions on the screen. Of the total observed, 2.4% of the neurons responded to a combination of spatial information and information about the object seen, in that they responded more the first time a particular image was seen in any position, or in other cases, only the first time that a particular stimulus had appeared in a particular position. The representation of space shown by these neurons was for the majority allocentric in that the spatial neuronal responses depended not on the position of the visual stimuli relative to the monkey's body axis, but instead on the position within the local allocentric frame of reference, that is the position of the visual stimulus on the visual display (Feigenbaum and Rolls, 1991). In rats, the responses of many hippocampal neurons depend on the rat's physical location (Leonard and McNaughton, 1990; O'Keefe, 1990, 1991), and it is suggested that part of the reason for this different representation of space is that in rats positions in space are most easily represented by places visited by the rats, whereas in primates positions in space can be defined without necessarily visiting them, but by remaining in one place and moving the eyes to look at different locations in space (Rolls, 1993; Rolls and O'Mara, 1993).

In another type of task for which the primate hippocampus is needed, conditional spatial response learning, in which the monkeys had to learn which spatial responses to make to different stimuli, (i.e., to acquire associations between visual stimuli and spatial responses), 14% of hippocampal neurons responded to particular combinations of stimuli and responses (Miyashita et al., 1989). Moreover, the activity of such neurons becomes modified during the learning of this task (Cahusac et al., 1993).

Thus not only is spatial information processed by the primate hippocampus, but it can be combined as shown by the responses of single neurons with information about which stimuli have been seen before (see Rolls, 1990b, 1993; Rolls et al., 1993). The ability of the hippocampus to form such arbitrary associations of information, probably originating from the parietal cortex about position in space with information originating from the temporal lobe about objects, may be important for its role in memory. Moreover these findings provide neurophysiological

support for the computational theory described here according to which such arbitrary associations should be formed onto single neurons in the hippocampus.

Another important type of information available from the hippocampal single neuron recordings is about the sparseness of the representation provided by hippocampal neurons. The sparseness, as defined below, indicates roughly the proportion of hippocampal neurons active at any one time in the hippocampus. The relevant time period for this measure is approximately 1s. Two arguments support this. First, the time course of hippocampal synaptic modification is a few hundred ms, in that the postsynaptic events involved in LTP operate over this time scale. Second, in normal environments in which stimuli might not change extremely rapidly, the stimuli forming an episode would be expected to be concurrent for a period of approximately 1s. In the monkey, in the different spatial and memory tasks in which recordings have been made, the evidence suggest that the sparseness is approximately 1%. For example, in the object and place memory task, 9% of cells responded in the task in which four places were to be remembered, so that approximately 2% of the neurons responded to any one place (Rolls et al., 1989). In a delayed spatial response task, 5.7% of the neurons had selective responses, and different neurons within this population responded to the different spatial responses required, in the sample, delay, and/or spatial response period, so that the proportion of neurons that responded in any one period in relation to any one spatial response was low (Cahusac et al., 1989). In rats, data from McNaughton's laboratory show that representations about place are very sparse in the dentate granule cells, are still sparse in CA3 neurons, and are fairly sparse in CA1 neurons (Barnes et al, 1990; Leonard and McNaughton, 1990; Jung and McNaughton, 1993). Estimates of these parameters are very important for defining how many memories could be stored in associative networks in the brain, as shown below and elsewhere (Rolls and Treves, 1990; Treves and Rolls, 1991).

THE CA3 NETWORK AS AN AUTOASSOCIATIVE MEMORY

Analysis of storage capacity

Our hypotheses concerning the operation of hippocampal circuits are based on the notion that the CA3 network operates as an autoassociative memory, that is, it is able, when provided with a small cue, to selectively retrieve a specific distribution of firing activity from among several stored on the same synaptic efficacies (strengths). This property is, in our view, intimately related to the architecture of the CA3 network, in particular to the prominent recurrent collateral connections. Others, such as Marr (1971), have noted that cued retrieval of one of a set of stored firing patterns can also be achieved with an alternative architecture, consisting of a cascade of feedforward associative nets (Willshaw et al., 1969), possibly assisted to some minor extent by the effect of recurrent collaterals (see Willshaw and Buckingham, 1990). We have shown, however, that the simple autoassociative architecture has at least two advantages over the multi-layered one: First, it requires many fewer synapses (by a factor equal to the number of layers needed to complete retrieval in the feedforward system; see Treves and Rolls, 1991) and, more importantly, it is possible to describe the conditions appropriate for learning, that is, the formation and storage of new patterns (Treves and Rolls, 1992), whereas no one has yet been able to describe how new firing patterns to be stored could be produced in the multi-layered associative net. In this section we briefly review our quantitative analyses of the storage and retrieval processes in the CA3 network (see Treves and Rolls, 1991, 1992).

To derive results applicable to CA3, we have extended previous formal models of autoassociative memory (Amit, 1989) by analyzing a network with graded response units to represent more realistically the continuously variable rates at which neurons fire, diluted connectivity, and sparse representations (Treves, 1990). We have found in general that 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

$$a = \frac{\langle r \rangle^2}{\langle r^2 \rangle} \tag{1}$$

where $\langle\ \rangle$ denotes an average over the statistical distribution characterizing the firing rate r of each cell in the stored patterns. Approximately,

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

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).

The main factors that determine the maximum number of memories that can be stored in an autassociative network are thus the number of connections on each neuron devoted to the recurrent collaterals, and the sparseness of the representation. For example, for $C^{RC} = 12{,}000$ and $a = 0.02$, p_{max} is calculated to be approximately 36,000. This storage capacity can be realised, with minimal interference between patterns, if the learning rule includes some form of heterosynaptic Long-Term Depression that counterbalances the effects of associative Long-Term Potentiation (Treves and Rolls, 1991).

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_t lost in retrieval and multiplying by p/C^{RC}:

$$I \equiv \frac{p}{C^{RC}} (i_p - i_t). \tag{3}$$

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.

The requirement of the input systems for efficient storage of new information

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 are dominated by the randomizing effect of the recurrent collaterals. This is the manifestation, in the CA3 network, of a general problem 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.

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, 1989a,b; Treves and Rolls, 1992). 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 EPSP). Second, the firing activity of granule cells appears to be very sparse (Jung and McNaughton, 1993) and this, together with the small number of connections per 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) 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 that would not happen with an occasionally firing fiber (Treves and Rolls, 1992).

Our argument based on this 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 perforant path input would not, the quantitative analysis shows, produce a pattern of firing in CA3 that contains sufficient information for learning (Treves and Rolls, 1992).

A different input system is needed to trigger retrieval

An autoassociative memory network needs afferent inputs also in the other mode of operation, that is, 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, that is, 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 an extensive 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,

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

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

Predictions arising from the analysis of CA3

Given the hypothesis, supported by the formal analysis above, that the mossy fiber system is particularly important dur-

[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 fibre inputs received by CA3 cells would follow a Poisson distribution of average 50/100 = 1/2, that is, 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 fibre 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.

ing storage of new information in the hippocampus, we have predicted that temporary inactivation of the mossy fibers should result, unless more generic effects complicate the picture, in amnesia specific to events occurring during the time window of inactivation, but not involving events learned before or after the inactivation. In contrast, inactivation of the perforant path to CA3, leaving intact, if possible, the initial part of the perforant path still connected to DG, should result in a relative deficit in learning new events, and also in a deficit in retrieving previously stored information. (This could be observed as an impairment in the performance of certain computationally demanding memory tasks of the type for which the hippocampus is required.) The deficit in learning would be due to the perforant path synapses not being appropriately modified during learning, thus rendering them unable to trigger retrieval from small cues. The deficit in retrieval would be due to the same synapses not transmitting during retrieval. This suggests experimental testing procedures and, interestingly, is consistent with experimental results already available (McNaughton et al., 1989). The roles we propose for the input systems to CA3 are also consistent with the general hypothesis (Rolls, 1987, 1989a–c, 1990a,b) that one of the functions of the processing occurring in the dentate gyrus is to transform certain neuronal representations into convenient inputs for the autoassociative memory we hypothesise is implemented in the CA3 network.

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, have also been shown to be computationally plausible based on computer simulations of the circuitry shown in the lower part of Figure 1. 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.

Dynamics and the temporal dimension

The arguments above have been developed while abstracting for a while from the time domain, by considering each pattern of activity, be it a new one or a memory being retrieved, as a steady-state pattern of firing rates. Reintroducing the temporal dimension poses a whole set of issues, on which the theoretical perspective has to confront available experimental evidence. For example, for how long are steady-state, or approximately steady-state, firing patterns sustained? Is this length of time compatible with the time requirements for associative synaptic plasticity as expressed via NMDA receptors? A related question is the following: 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? How does the time scale for recurrent activation compare, for example, with the period of the theta rhythm in the rat?

A partial answer to these last questions can be inferred from recent theoretical developments based on the analysis of the collective dynamic properties of realistically modeled neuronal units (Treves, 1993; Treves et al., 1994). The analysis indicates that the evolution of distributions of activity in a recurrent network follows a plurality of different time scales; and that the most important of those time scales are only mildly dependent on single-cell properties, such as prevailing firing rates or membrane time constants, but rather depend crucially on the time constants governing synaptic conductances. This result suggests that in the case of CA3, the activation of recurrent collaterals between pyramidal cells, through synapses whose time constant is rather short (a few ms), will contribute to determining the overall firing pattern within a period of a very few tens of ms (see Treves, 1993; Treves et al., 1994, for further details). (In animals in which there is a hippocampal theta rhythm, such as the rat, the relevant CA3 dynamics could take place within a theta period. In monkeys, no prominent theta rhythm has been found in the hippocampus.) The indication is thus that retrieval would be very rapid from the CA3 network, indeed, fast enough for it to be biologically plausible.

THE DENTATE GRANULE CELLS

The theory is developed elsewhere that the dentate granule cell stage of hippocampal processing that precedes the CA3 stage acts in four ways to produce during learning the sparse yet efficient (i.e., non redundant) representation in CA3 neurons which is required for the autoassociation to perform well (Rolls, 1989a–c, 1993; 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 matrix 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:

$$\delta w_{ij} = k \cdot r_i (r_j' - w_{ij}), \quad (6)$$

where 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 long-term depression as well as long-term potentiation (see Levy and Desmond, 1985; Levy et al. 1990)

makes the sum of the synaptic weights on each neuron remain roughly constant during learning (see Oja, 1982; Rolls, 1989c).

The second way is also a result of the competitive learning hypothesized to be implemented by the dentate granule cells (Rolls, 1987, 1989–c, 1990a,b, 1994). It is proposed that this allows overlapping inputs to the hippocampus to be separated, in the following way (see also Rolls, 1994). 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 perfectly, as has been shown elsewhere (Rolls, 1989c). 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, and could enable it to solve what have been called configural types of learning task (Sutherland and Rudy, 1991). (We know that it is a separate and 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. Our view is that such black-and-white 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; one might just expect it to be better with the hippocampus.)

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

A fourth way is that as suggested and explained above, 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.

IS THE CA1 ORGANIZATION DETERMINED BY COMPUTATIONAL CONSTRAINTS?

An information storage device or a pointer?

In applying quantitative arguments to analyze CA3 organization, we have adopted an approach based on the notion of information. Before using the same approach to investigate, in turn, CA1, it is important to note the full implications of the approach for theories of hippocampal function. On a general level, this approach shares with others the assumption that a role of the hippocampus is to record the unique combination of elements (e.g., ABCDE) that, having been associated together at one time by temporal contiguity, make up the memory of an episode or event. In the approach that we have followed here, it is assumed that the hippocampus is, temporarily, the actual storage site of the episodic memory. In the time span before consolidation of the memory occurs in neocortex, the hippocampus is taken to be responsible for holding and being able to produce all the data about the episode, within the limits set by its anatomical resources and the concurrent storage of as many other episodes as possible. A different approach, however, is to hypothesize that the hippocampus stores only pointers, in which case it would not need to retrieve all the components of an episode, but only a pointer to identify that episode from among those episodes whose pointers are stored in the hippocampus. According to this prescription, the hippocampus would output a pointer during recall, and the components of the episode would have to be retrieved in other brain structures using the pointer. (To make the difference explicit, let us assume that we need to store 36,000 memories in the hippocampus, each needing 100 bits of information to encode. If the components of the episode are stored in the hippocampus, the amount of storage required in the hippocampus is 3,600,000 bits. If pointers are stored, the amount of information to be stored in the hippocampus is $\log_2 36{,}000 \sim 15$ bits/memory, or 540,000 bits for 36,000 memories. If the pointer hypothesis is modified to assume that the hippocampus must store pointers that identify recent episodic memories from a larger number (say 10^6) of stored memories, which includes all those previously stored, then the hippocampus would need to store $\log_2 10^6 = 20$ bits/memory, which is a very small amount of information compared to the capacity of the CA3 network.) According to the pointer hypothesis, the hippocampus would output the pointer to a stored memory (e.g., its serial number), and a cortical area such as area 7 might retrieve its part of the information using the pointer.

Which of the above viewpoints better describes hippocampal function is a matter that, of course, will be ultimately decided by experimental data. It is nevertheless important to analyze the implications of each hypothesis in detail. One of the uses of quantitative models such as those we propose here is precisely that of making explicit the predictions and possibilities for falsification, associated with each theoretical view.[2]

Preserving the information content of CA3 patterns

The amount of information about each episode retrievable from CA3 has to be balanced, as we have said, against the number of episodes that can be held concurrently in storage.

[2]One relevant number here is that we have estimated that the maximum amount of information that could be stored and recalled from the CA3 network in the rat, with 12,000 recurrent collateral synapses and a sparseness of encoding a that allows three times as many memories to be stored as there are recurrent collaterals, is as high as 1,400 Mbits (under the ideal case assumption that different cells code for independent information: the extent to which this will be reduced by non-independence of neuronal activity in the brain is a topic of recent investigation; see Treves and Rolls, 1991).

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 the 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 propose here that the need to preserve the full information content present in the output of an autoassociative memory requires an intermediate recording 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 non modifiable, 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. In addition, if the total amount of information carried by CA3 cells is redistributed over a larger number of CA1 cells, less information can be loaded onto each CA1 cell, rendering the code more robust to information loss in the next stages. If each CA3 cell had to code for 2 bits of information, for example, 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. Minimizing information loss also turns out to favor, but only very marginally, a recoding stage that performs a very uniform sampling of the autoassociative output, with each CA1 unit receiving approximately equal numbers of synaptic contacts from the CA3 cells. (Details of this calculation will be presented elsewhere.)

Recoding in CA1

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, each part—A, B, C, D and E of a particular episode—would 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 sparse.

The perforant path projection to CA1

One major 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 definite proportion (about ⅓; Amaral et al., 1990) from the 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. The existence of the double set of inputs to CA1 appears to indicate, in general, that there is a need for a system of projections that makes again available, at the output of CA3, information closely related to that which was originally given as input to the hippocampal formation.

An approach based on a quantitative analysis suggests an explanation for this. During retrieval, the information content of the firing pattern produced by the hypothesized CA3 autoassociative net is invariably reduced both with respect to that originally present during learning, and, even more, with respect to that available, during learning, at the input to CA3. The cue that, transmitted through entorhinal cortex, is used to elicit retrieval of the full event in CA3, may instead carry rather detailed information on a limited number of elements of the episode. The perforant path projection might thus serve, during retrieval, to integrate the information-reduced description of the full event recalled from CA3 with the information-richer description of only those elements used as a cue provided by the entorhinal/perforant path signal. This integration may be useful both in the consolidation of longer-term storage and in the immediate use of the retrieved memory.

BACKPROJECTIONS TO THE NEOCORTEX

The hippocampus as a buffer store

The concept we advocate is that the hippocampus is an intermediate buffer store, into which information is stored in real time as events happen in the world. Helped by the Hebb-like learning rule implemented in networks in many parts of the hippocampus, especially the CA3 recurrent collateral network, the hippocampus is able to rapidly store "snapshots" or

episodes, where the minimum duration of each episode is determined by the period over which the Hebbian associativity operates as shown by LTP, approximately 1s. Such an episodic memory would be useful in solving a number of spatial and nonspatial tasks that require the hippocampus (see summary in the first section and Rolls, 1990a–c, 1991; Rolls and O'Mara, 1993). Within each episode, there would be insufficient time to reorganize the information, so it would necessarily be stored in the form in which it arrived, as a "snapshot." It is, of course, likely that a new snapshot is not stored every second. Instead, it is likely that episodes are stored, particularly when something new happens in the environment, or when something reinforcing happens in the environment, such as an emotional or motivational event (Rolls, 1990a–c). Some of the nonspecific input systems to the hippocampus, such as the cholinergic input from the medial septum—the activity of which is influenced by brainstem arousal systems—may perform this function by increasing activation in the hippocampus, thus helping some neurons to exceed the threshold for synaptic modification induced by activation of NMDA receptors, and thus enabling the storage of information in the synaptic matrix (Rolls, 1989b).

Two questions that arise from this are considered next. One is the period over which the hippocampal buffer store would be useful, and the implications of this for understanding retrograde amnesia produced by damage to the hippocampal system. The other is how the hippocampal buffer store could make its contents available to the neocortex.

Capacity limitations of the buffer store, and the gradient of retrograde amnesia

As shown above, the number of memories that can be stored in the hippocampal CA3 network is limited by the number of synapses on the dendrite of each CA3 cell devoted to CA3 recurrent collaterals (12,000 in the rat), increased by a factor of perhaps 3–5 depending on the sparseness of activity in the representations stored in CA3. Let us assume for illustration that this number is 50,000. If more than this number were stored, then the network would become unable to retrieve memories correctly. Therefore, there should be a mechanism that prevents this capacity being exceeded, by gradually overwriting older memories. This overwriting mechanism is in principle distinct from the heterosynaptic depression implicit in the covariance learning rule hypothesized above to be present in the hippocampus. The heterosynaptic depression is required in order to minimize interference between memories held at any one time in the store, rather than in order to gradually delete older memories. A variety of formulas, representing modifications of the basic covariance rule, may be proposed in order to explicitly model the overwriting mechanism. Whatever detailed model one assumes, the important statement about such an overwriting process is that its pace should be determined by the rate at which new memories are to be stored in the hippocampal buffer, and by the need to keep the total number concurrently stored below the capacity limit. The average time after which older memories are not retrievable any more from the hippocampal store will not, therefore, be a characteristic time per se, but rather the time in which a number of new episodes, roughly of the order of the storage capacity, has been acquired for new storage in the hippocampal system.

Given the estimates for the number of memories that can be stored in the CA3 network described above, one can link the estimated number p with a measure of the time gradient of retrograde amnesia. Such measures have been tentatively produced recently for humans, monkeys, and rats (Squire, 1992), although there is still discussion about the evidence for a temporal gradient for retrograde amnesia (Gaffan, 1993). To the extent that there may be such a gradient as a function of the remoteness of the memory at the onset of amnesia, it is of interest to consider a possible link between the gradient and the number of memories that can be stored in the hippocampus. The notion behind the link 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, for example, $p = 50{,}000$ (somewhat larger than a typical estimate for the rat, if C is larger and the sparseness perhaps smaller), 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/s, as noted above.)

The main point just made is that the hippocampus would be a useful store for only a limited period, which might be 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 buffer store, the representations in the buffer store would remain stable and persistent. Our hypotheses have clear experimental implications, both for recordings from single neurons and for the gradient of retrograde amnesia, in stable or frequently changing environments.

This analysis makes it clear that the memories held in the hippocampal buffer store should be stored elsewhere as time progresses, if they are needed in the long term, and if many new episodic memories are being acquired, as considered next.

The transfer of information from the hippocampus to the cerebral cortex

We have seen that the hippocampus appears to have an appropriate architecture to act as a buffer store for unprocessed episodic memories, stored in real time as events happen in the world. This may be contrasted with the type of storage that is needed for long-term memory. In long-term memory storage, much more organization of the information is likely to be needed. This is especially true for semantic memories (McClelland et al., 1992). For example, a series of episodic memories stored during a particular journey from A to E via B, C, and D might contain information that could usefully be

added to a semantic network which contained geographical information about the relations between some but not all of the places visited on the particular journey. The important constraint here is to add information to an existing network without overwriting or disrupting its existing contents. This is probably best performed by making a succession of small adjustments to the existing semantic network (McClelland et al., 1992). In any case the process is very different from that involved in forming an episodic memory "on the fly," in which the whole collection of subevents present in the episode must be stored together strongly, and must be kept separate from other episodes, even if the episodes are somewhat similar.

Although this discussion has been in terms of adding to long-term semantic memories, it is envisaged that the long-term storage of episodic memories would also involve adding to a neocortical episodic memory store, for which some reorganization of the hippocampal episodic memories would be appropriate. For example, in the long-term episodic store, it could be important not just to have a very large collection of unassorted episodic memories, but to include with them information about the order in which they occurred, and to perhaps link those concerned with similar themes (e.g., memories of birthdays).

It has also been noted elsewhere (O'Kane and Treves, 1992) that a system that attempted to store episodic memories as such in the neocortex would be constrained by the same capacity limit which applies to the CA3 network, that is, a limit determined by the number of synaptic connections per cell, not by the number of cells. Some form of reorganization of episodic information appears therefore to be necessary in order to make efficient use of the much larger storage space available in the neocortex.

What is suggested is that the hippocampus is able to recall the whole of a previously stored episode for a period of days, weeks or months after the episode, when even a fragment of the episode is available to start the recall. This recall from a fragment of the original episode would take place particularly as a result of completion produced by the autoassociation implemented in the CA3 network. It would then be the role of the hippocampus to reinstate in the cerebral neocortex (via the subiculum and entorhinal cortex) the whole of the episodic memory. The cerebral cortex would then, with the whole of the information in the episode now producing firing in the correct sets of neocortical neurons, be in a position to incorporate the information in the episode into its long-term store in the neocortex. The operation of this semantic storage in the neocortex may benefit from serial cognitive processes that can focus attention on the links that must be made from the new pieces of episodic information into the existing semantic store, but is not a process we consider further here. What we do consider further here is how recall of the whole episode within the hippocampus could lead to reinstatement of the information about the episode back in the cerebral neocortex.

We suggest 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 subicular neurons, which in turn would 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 the 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 the 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 related 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).

Our understanding of the architecture with which this would be achieved is shown in Fig. 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 the entorhinal cortex would be associatively modified. A similar process would be implemented at preceding stages of the 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–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-milliseconds 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 deaddressing 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 exactly how 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 the apical dendrites of cortical pyramidal cells.

If the backprojection synapses are associatively modifiable, we may consider the duration of the period for which their synaptic modification should persist. What 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 (see Kirkwood et al., 1993). It is noted that it would be appropriate to have this similarity of temporal time course for at least one stage in the series of backprojecting stages from the CA3 region to the neocortex. Such stages might include the CA1 region, subiculum, entorhinal cortex, and perhaps the parahippocampal gyrus. However, from the multimodal cortex (e.g., the parahippocampal gyrus) back to earlier cortical stages, it might be desirable for the backprojecting synapses to persist for a long period, so that some types of recall and top-down processing (see Rolls, 1989a,b) mediated by the operation of neocortical–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., 1992). 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 recent memories stored in the hippocampus might be recalled in regions of the cerebral neocortex.

Quantitative constraints on the connectivity

How many backprojecting fibers does one need to synapse on any given neocortical pyramidal cell, in order to implement the mechanism outlined above? Clearly, if the theory were to produce a definite constraint of that sort, quantitative anatomical data could be used for verification or falsification.

Attempts to come up with an estimate of the number of synapses required have sometimes followed the simple line of reasoning presented next. Consider first the assumption that hippocampal–cortical connections are monosynaptic and not modifiable, and that all existing synapses have the same efficacy. Consider further the assumption that a hippocampal representation across N cells of an event consists of $a_h N$ cells firing at the same elevated rate, while the remaining $(1-a_h)N$ cells are silent. If each pyramidal cell in the association areas of the neocortex receives synapses from an average C^h hippocampal axons, there will be an average probability $y = C^h/N$ of finding a synapse from any given hippocampal cell to any given neocortical one. Across neocortical cells, the number A of synapses of hippocampal origin activated by the retrieval of a particular episodic memory will follow a Poisson distribution of average $ya_hN = a_hC^h$:

$$P(A)=(a_hC^h)^A \exp(-a_hC^h)/A!. \quad (7)$$

The neocortical cells activated as a result will be those, on the tail of the distribution, receiving at least T active input lines, where T is a given threshold for activation. Requiring that $P(A>T)$ be at least equal to a_{nc}, the fraction of neocortical cells involved in the neocortical representation of the episode, results in a constraint on C^h. This simple type of calculation can be extended to the case in which hippocampal-cortical projections are taken to be polysynaptic, and mediated by modifiable synapses. In any case, the procedure does not appear to produce a very meaningful constraint for at least three

reasons. First, the resulting minimum value of C^h, being extracted by looking at the tail of an exponential distribution, varies dramatically with any variation in the assumed values of the parameters a_h, a_{nc}, and T. Second, those parameters are ill-defined in principle: a_h and a_{nc} are used having in mind the unrealistic assumption of binary distributions of activity in both the hippocampus and neocortex (although the sparseness of a representation can be defined in general, as shown above, it is the particular definition pertaining to the binary case that is invoked here); while the definition of T is based on unrealistic assumptions about neuronal dynamics (how coincident does one require the various inputs to be in time in order to generate a single spike, or a train of spikes of given frequency, in the postsynaptic cell?). Third, the calculation assumes that neocortical cells receive no other inputs, excitatory or inhibitory. Relaxing this assumption to include, for example, nonspecific activation by subcortical afferents, makes the calculation impossible to start with.

An alternative way to estimate a constraint on C^h, still based on very simple assumptions, but hopefully producing a result which is more robust with respect to relaxing those assumptions, is the following. 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 backwards toward the neocortex, remains 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 Eq. (2):

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

where, however, a is now the sparseness of the representation at any given stage, and C is the average number of (backprojections) 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 buffer, 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 $C^{HBP} = C^{RC}\, a_{nc}/a_{CA3}$, 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 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–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 the number 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 (see 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 that might possibly originate in the amygdala (see Amaral, 1992) or in multimodal cortical areas (allowing, e.g., 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 that 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. 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 that forms episodic memories which could be made available locally within an area of the cerebral neocortex.

DISCUSSION

Comparison with other theories of hippocampal function

We have produced hypotheses on 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 we highlight the differences in the following section, we note that the overall view we have 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 we do not, of course, claim priority on all the propositions put forward in this report.

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 that 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) 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 our theory, in which the hippocampus is an intermediate storage memory, 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 a local view is seen later, 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, not autoassociation, and it has been postulated that this pattern association can be performed by the hippocampus (see McNaughton and Morris, 1989). This theory is again in contrast to our theory, in which the hippocampus is an intermediate storage memory utilizing autoassociation, which is able to recall what was stored in it, using a partial cue if necessary. Our theory is not inconsistent with the presence of view cells and whole body motion (e.g., vestibular) cells in the primate hippocampus (Rolls and O'Mara, 1993; or local view cells in the rat hippocampus and head direction cells in the rat presubiculum), for we note that 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.

Our theory thus holds that the hippocampus is used for the formation of episodic memories, by acting as an intermediate-term autoassociative memory. 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, if necessary, recalled from the hippocampus). 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 can lead to problems such as topographical and other spatial agnosias (see Kolb and Whishaw, 1990). This is consistent with spatial computations normally being performed in the parietal cortex. (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). (We have noted elsewhere the formal similarity of a feedforward multistage associative memory with no recurrent processing to a single stage recurrently connected autoassociative network—Treves and Rolls, 1991. The number of stages required can be thought of as being equivalent to the number of iterations used in a one-layer recurrent network. The total number of synapses required in a multistage network is likely to be much larger, and training it with a local learning rule is not plausible.) The technical approach Marr took to the analysis was based on probabilities computed in the tail of Poisson distributions, and as we noted above, 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. 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, e.g., 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 dream sleep. We, on the other hand, postulate that in order for the episodic memories to be incorporated into long-term memories, which will often be semantic and may require small modifications to links 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. 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 we are aware ever produce such a theory. In this report, we 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. We identify this system with the multistage backprojection system from the hippocampus to the cerebral neocortex, and between adjacent neocortical areas.

Another theory is that the hippocampus (and amygdala) are 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). We take the approach 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 from 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. We therefore predict 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. Thus, we believe that what is found after hippocampal damage will reflect quantitatively rather than just qualitatively the ability to form new (especially multimodal, with one modality space) episodic memories. We also note that Mishkin's theory was a theory of what the hippocampus does, whereas ours is a theory of what the hippocampus does and how it does it.

Another theory of hippocampal function is that it is necessary for the formation of configurative memories (Sutherland and Rudy, 1991), that is when a memory must be separated from memories of subsets of its elements. Our theory places emphasis on the autoassociation capability of the CA3 system, and it is the case that autoassociation memories cannot resolve the subset problem, because the subset of information will always be taken as a partial cue for the larger memory that includes that subset. However, an important aspect of our theory of hippocampal function is that the CA3 system is preceded by the dentate system which acts as a competitive network. This is able to produce separate outputs for subsets of events from that to a whole set of events, and ensures that different patterns reach the CA3 cells via the mossy fibers during learning, so that different representations are produced and stored in the CA3 autoassociation system for subsets from the whole set of events. By making such overlapping patterns very different from each other, the dentate system enables the hippocampal CA3 system to store different memories for even somewhat similar episodes, a very important feature of an episodic memory system. Insofar as the hippocampal system has circuitry specialized for this in the dentate granule cells, then we believe that this could allow the hippocampus to make a special contribution to the storage of such overlapping memories, even though as noted above, this is likely to be measurable as a quantitative advantage in memory tasks of having a hippocampus, rather than an all-or-none consequence of having a hippocampus or not.

Another suggestion on the function of the hippocampus is that it stores pointers to memories located in the neocortex (Teyler and Discenna, 1986). We have noted that one argument which suggests that this may not be the case is that the potential storage capacity of the hippocampus is much larger than would be required to store only pointers to neocortical memories. The implication is that the different subevents that must be remembered as an episodic memory are represented in the hippocampus and linked together by hippocampal association mechanisms to form the episodic memory (see Treves and Rolls, 1992). Another contrast is that the pointer theory must assume that when episodic memories are formed, they are formed on-line in the cerebral neocortex, with a pointer being simultaneously set up in the hippocampus. We have noted above that storage of episodic memories on-line in a permanent (neocortical) memory store is unlikely because more organization of the memory storage, including appropriate modifications made to existing memory structures, is needed than would be possible in real time. Further, because the number of episodic memories that could be stored without reorganization in the neocortex is limited by the number of synapses per cell not by the number of cells, just as in the hippocampus, even the whole of the neocortex would offer no better advantage for the storage of episodic memories than is offered by the hippocampus (see O'Kane and Treves, 1992). Moreover it is the case that many episodic events need not be allocated permanent storage, so that initial storage of episodic memories in an intermediate buffer store, such as appears to be provided by the hippocampus, is advantageous.

Another theory is being developed by McClelland et al. (1992). This is still at the systems level, and has not yet explored quantitatively how the system would operate. It is similar to

our theory at the systems level, except that the last set of synapses that are modified rapidly during the learning of each episode are those between the CA3 and the CA1 pyramidal cells (see Fig. 1). 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 the perforant path projection to CA3) must be associatively modifiable if recall is to operate efficiently (see Treves and Rolls, 1992). Our view, therefore, is that it is likely that at least as far back into neocortical processing as the inferior temporal cortex, the backprojecting synapses should be associatively modifiable, as quickly as 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.

The aim of this comparison of our 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. We recapitulate here the main predictions arising from our theory, all of which could 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, with a similar time course to that of associative modifiability within the hippocampus itself.
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, our quantitative analyses predict 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.

CONCLUSION

We have drawn together a number of recent neurophysiological, neuroanatomical, and theoretical investigations, and shown how they fit 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. We have shown, in more detail than have previous reports, how hippocampal recall of recent memories could via backprojections to the cerebral cortex lead to recall in the cerebral cortex, and suggested how this could be useful to the cerebral cortex in forming new long-term memories. We have also suggested a number of experimental tests of the theory. The results presented here also start to lead toward a quantitative understanding of the gradient of retrograde amnesia produced by damage to the hippocampus and related structures.

ACKNOWLEDGMENTS

Different parts of the research described here were supported by the Medical Research Council (PG8513790), by an EEC BRAIN grant, by the MRC Oxford Research Centre in Brain and Behaviour, by the Oxford McDonnell-Pew Centre in Cognitive Neuroscience, and by a Human Frontier Science program grant. We thank S.O'Mara and P.Foldiak for comments on an earlier version of this paper. A.T. also acknowledges support from the INFM and INFN of Italy.

REFERENCES

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

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

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

Barnes CA, McNaughton BL, Mizumori SJ, Lim LH (1990) Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. Prog Brain Res 83:287–300.

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.

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.

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 hippocampal neurons in the monkey during the learning of a conditional spatial response task. Hippocampus 3:29–42.

Churchland PS, Sejnowski TJ (1992) The Computational Brain. Cambridge, MA: MIT Press.

Collingridge GL, Singer W (1990) Excitatory amino acid receptors and synaptic plasticity. Trends Pharmacol Sci 11:290–296.

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, Harrison S (1989) Place memory and Scene memory: effects of fornix transection in the monkey. Exp Brain Res 74:202–212.

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. Behavi 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.

Insausti R, Amaral DG, Cowan WM (1987) The entorhinal cortex of the monkey. II. Cortical afferents. J Comp Neurol 264:356–395.

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.

Jung MW, McNaughton BL (1993) 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.

Kubie JL, Muller RU (1991) Multiple representations in the hippocampus. Hippocampus 1:240–242.

Leonard BW, McNaughton BL (1990) Spatial representation in the rat: conceptual, behavioral and neurophysiological perspectives. In: Neurobiology of comparative cognition (Kesner RP, Olton DS, eds) pp 363–422. Hillsdale, NJ: Erlbaum.

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

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

McClelland JL, McNaughton BL, O'Reilly R, Nadel L (1992) Complementary roles of hippocampus and neocortex in learning and memory. Soc Neurosci Abstr 18:1216.

McNaughton BL, Morris RGM (1989) Hippocampal synaptic enhancement and information storage within a distributed memory system. Trends Neurosci 10:408–415.

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, 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, 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.

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

Miles R (1988) Plasticity of recurrent excitatory synapses between CA3 hippocampal pyramidal cells. Soc for Neurosci Abstr 14:19.

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 Soc Lond [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- [^{3}H]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, Ch. 11 (Morris RGM, ed), 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'Keefe J (1991) The hippocampal cognitive map and navigational strategies. In: Brain and space, Ch. 16 (Paillard J, ed), pp 273–295. Oxford: 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, Ch. 13 (Byrne JH, Berry WO, eds), 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, Ch. 8 (Durbin R, Miall C, Mitchison G, eds), 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, Ch. 12 (Olton DS, Kesner RP, eds), pp 339–362. Hillsdale, NJ: Erlbaum.

Rolls ET (1990c) A theory of emotion, and its application to understanding the neural basis of emotion. Cognition and 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: neurobiological aspects of emotion, memory, and mental dysfunction, Ch. 5 (Aggleton JP, ed), 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 Lond [Biol] Soc 335:11–21.

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

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 (1990) The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. Network 1:407–421.

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, ch. 17 (Ono T, Squire LR, Raichle M, Perrett D, Fukuda M, eds), pp 276–300. New York: Oxford University Press.

Rolls ET, Cahusac PMB, Feigenbaum JD, Miyashita Y (1993) Responses of single neurons in the hippocampus of the macaque related to recognition memory. Exp Brain Res 93:299–306.

Rupniak NMJ, Gaffan D (1987) Monkey hippocampus and learning about spatially directed movements. J Neurosci 7:233–2337.

Seress L (1988) Interspecies comparison of the hippocampal formation shows increased emphasis on the regio superior in the Ammon's horn of the human brain. J Hirnforsch 29:335–340.

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 Ch. 12 (Byrne J, Berry WO, eds), pp 208–239. New York: Academic Press.

Storm-Mathiesen J, Zimmer J, Ottersen OP (editors) (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.

Teyler TJ, Discenna P (1986) The hippocampal memory indexing theory. Behav Neurosci 100:147.

Treves A (1990) Graded-response neurons and information encodings in autoassociative memories. Physical Rev A 42:2418–2430.

Treves A (1993) Mean-field analysis of neuronal spike dynamics. Network 4:259–284.

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, Tovee MJ (1994) On the time required for recurrent processing in the brain.

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

West MJ, Gundersen HGJ (1990) Unbiased stereological estimation of the numbers of neurons in the human hippocampus. J Comp Neurol 296:1–22.

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.

Willshaw DJ, Buneman OP, Longuet-Higgins AC (1969) Non-holographic associative memory. Nature 222:960–962.

Wilson FAW, Rolls ET (1990a) Neuronal responses related to the novelty and familiarity of visual stimuli in the substantia innominata, diagonal band of Broca and periventricular region of the primate. Exp Brain Res 80:104–120.

Wilson FAW, Rolls ET (1990b) Neuronal responses related to reinforcement in the primate basal forebrain. Brain Res 502:213–231.

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.

Zola-Morgan S, Squire LR, Amaral DG (1986) Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. J Neurosci 6:2950–2957.

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