A computational theory of episodic memory formation in the hippocampus

Edmund T. Rolls*,1
Oxford Centre for Computational Neuroscience, Oxford, United Kingdom

Abstract

A quantitative computational theory of the operation of the hippocampus as an episodic memory system is described. The CA3 system operates as a single attractor or autoassociation network to enable rapid, one-trial associations between any spatial location (place in rodents or spatial view in primates) and an object or reward and to provide for completion of the whole memory during recall from any part. The theory is extended to associations between time and object or reward to implement temporal order memory, also important in episodic memory. The dentate gyrus performs pattern separation by competitive learning to produce sparse representations, producing for example neurons with place-like fields from entorhinal cortex grid cells. The dentate granule cells produce by the very small number of mossy fibre connections to CA3 a randomizing pattern separation effect important during learning but not recall that separates out the patterns represented by CA3 firing to be very different from each other, which is optimal for an unstructured episodic memory system in which each memory must be kept distinct from other memories. The direct perforant path input to CA3 is quantitatively appropriate to provide the cue for recall in CA3, but not for learning. The CA1 recodes information from CA3 to set up associatively learned backprojections to neocortex to allow subsequent retrieval of information to neocortex, providing a quantitative account of the large number of hippocampo-neocortical and neocortical–neocortical backprojections. Tests of the theory including hippocampal subregion analyses and hippocampal NMDA receptor knockouts are described and support the theory.

© 2010 Elsevier B.V. All rights reserved.
1. Introduction

In this paper a computational theory of how the hippocampus operates in episodic memory is described. I consider how the network architecture of the hippocampal system may enable it to implement episodic memory, and to recall the whole of a memory in the neocortex when a partial retrieval cue is present. I focus on a fundamental property of episodic memory, the ability to store and retrieve the memory of a particular single event involving an association between items such as the place and the object or reward seen at that place. Episodic memory in the sense of a series of linked events requires this type of event memory, and could be implemented by linking together a series of events. After the theory is presented, I then describe neurophysiological and related evidence that tests the theory and provides further evidence on how memory is implemented in the hippocampus.

Episodic memory, the memory of a particular episode, requires the ability to remember particular events and to distinguish them from other events. An event consists of a set of items that occur together, such as seeing a particular object or person’s face in a particular place. An everyday example might be remembering where one was for dinner, who was present, what was eaten, what was rewarded when the object is seen alone. The ability to recall a whole memory when the place at which an object was shown must be recalled. A related task is whole scene memory, in which a monkey must learn which locations in each of a number of scenes if touched lead to reward. The task can be thought of as one in which a place in a scene and reward must be associated together by learning. In rodents, the watermaze [98] and cheeseboard [141] tasks can be thought of as testing reward-place association memory.

2. Systems-level functions of the primate hippocampus

Any theory of the hippocampus must state at the systems level what is computed by the hippocampus. Some of the relevant evidence comes from the effects of damage to the hippocampus, the responses of neurons in the hippocampus during behavior, and the systems-level connections of the hippocampus, described in more detail elsewhere [141,145]. As described in Section 1, evidence of spatial representations in the hippocampus, and of whether these spatial representations can be combined with object representations in memory tasks such as object-place memory, is relevant to understanding the functions of the primate hippocampus in episodic memory. In object–place memory, the place where an object was seen must be remembered. A related task is whole scene memory, in which a monkey must learn which locations in each of a number of scenes if touched lead to reward. The task can be thought of as one in which a place in a scene and reward must be associated together by learning. In rodents, the watermaze [98] and cheeseboard [141] tasks can be thought of as testing reward-place association memory.

2.1. Evidence from the effects of damage to the primate hippocampus

Damage to the hippocampus or to some of its connections (described in Sections 2.2 and 3.1) such as the fornix in monkeys produces deficits in learning about the places of objects and about the places where responses should be made [17]. For example, macaques and humans with damage to the hippocampal system or fornix are impaired in object–place memory tasks in which not only the objects seen, but where they were seen, must be remembered [19,25,40,43,112,154]. Posterior parahippocampal lesions in macaques impair even a simple type of object-place learning in which the memory load is just one pair of trial-unique stimuli [85]. It is further predicted that a more difficult object–place learning task with non-trial-unique stimuli and with many object–place pairs would be impaired by neurotoxic hippocampal lesions.

Further, neurotoxic lesions that selectively damage the primate hippocampus impair spatial scene memory, tested by the ability to remember where in a scene to touch to obtain reward [104]. Also, fornix lesions impair conditional left–right discrimination learning, in which the visual appearance of an object specifies whether a response is to be made to the left or the right [150]. A comparable deficit is found in humans [113]. Fornix sectioned monkeys are also impaired in learning on the basis of a spatial cue which object to choose (e.g. if two objects are on the left, choose object A, but if the two objects are on the right, choose object B) [42]. Monkeys with fornix damage are also impaired in using information about their place in an environment. For example, there are learning impairments when which of two or more objects the monkey had to choose depended on the position of the monkey in the room [41]. More recently, Banta Lavenex and Lavenex have described deficits produced by hippocampal damage in monkeys performing allocentric spatial memory tasks [12]. One such task involved freely moving in an environment using allocentric spatial room cues to remember the locations of inverted cups that contained food. This is a food reward–allocentric place association task. Rats with hippocampal lesions are also impaired in using environmental spatial cues to remember particular places [22,61,68,87,109], to utilize spatial cues or to bridge delays [65,66,68,116,141], or to perform relational operations on remembered material [34]. In humans, functional neuroimaging shows that the hippocampal system is activated by allocentric spatial processing [20,52].
Many of these memory functions are important in event or episodic memory, in which the ability to remember what happened where on typically a single occasion (or trial in a learning experiment) is important. It will be suggested below that an autoassociation memory implemented by the CA3 neurons enables event or episodic memories to be formed by enabling associations to be formed between spatial and other including object or reward representations.

Information stored in the hippocampus will need to be retrieved and affect other parts of the brain in order to be used. The information about episodic events recalled from the hippocampus could be used to help form semantic memories [119,120,124,171]. For example, remembering many particular journeys could help to build a geographic cognitive map in the neocortex. The hippocampus and neocortex would thus be complementary memory systems, with the hippocampus being used for rapid, “on the fly”, unstructured storage of information involving activity potentially arriving from many areas of the neocortex; while the neocortex would gradually build and adjust on the basis of much accumulating information, often recalled from the hippocampal unstructured store, the semantic representation [88,102,120,171]. The theory described below shows how information could be retrieved within the hippocampus, and how this retrieved information could enable the activity in neocortical areas that was present during the original storage of the episodic event to be reinstated, thus implementing recall, using hippocampo-neocortical backprojections (see Fig. 1).

### 2.2. Systems-level anatomy

To understand the functions of the primate hippocampus in event or episodic memory, it is necessary to understand which other parts of the brain it receives information from. Does it for example receive object as well as spatial information in terms of its connectivity? The primate hippocampus receives inputs via the entorhinal cortex (area 28) and the highly developed parahippocampal gyrus (areas TF and TH) as well as the perirhinal cortex from the ends of many processing streams of the cerebral association cortex, including the visual and auditory temporal lobe association cortical areas, the prefrontal cortex, and the parietal cortex [3,6,76,141,145,164,175,180] (see Fig. 1). The hippocampus is thus by its connections potentially able to associate together object and spatial representations. In addition, the entorhinal cortex receives inputs from the amygdala, and the orbitofrontal cortex, which could provide reward-related information to the hippocampus [21,114,155,163].
The primary output from the hippocampus to neocortex originates in CA1 and projects to subiculum, entorhinal cortex, and parahippocampal structures (areas TF–TH) as well as prefrontal cortex [32,174,175,178] (see Fig. 1), though there are other outputs [141]. These are the pathways that are likely to be involved in the recall of information from the hippocampus.

The systems-level analysis of neuronal activity in the hippocampus provides tests of the theory of hippocampal function in episodic memory and how this is implemented, and this is described in Section 4, and other evidence that tests the theory is described in Section 5.

3. A theory of the operation of hippocampal circuitry as a memory system

In this section, I consider how event or episodic memories might be retrieved by hippocampal circuitry, and in addition retrieved back into the neocortex. The theory has been developed through many stages [118–121,123,124,126,128,129,133,141,145,149,169–171], has as a predecessor developments made by David Marr [86] who participated with me in lectures on the hippocampus by L. Weiskrantz at Cambridge, and has benefitted greatly from collaborations with many whose names appear below in the citations, including Alessandro Treves and Simon Stringer.

3.1. Hippocampal circuitry Fig. 1 [4,7,11,72,76,105,156,180,181]

[see Fig. 1 and 4, 7, 11, 72, 76, 105, 156, 180, 181]

Projections from the entorhinal cortex layer 2 reach the granule cells (of which there are 10^6 in the rat) in the dentate gyrus (DG), via the perforant path (pp) [178]. The granule cells project to CA3 cells via the mossy fibres (mf), which provide a sparse but possibly powerful connection to the 3 × 10^5 CA3 pyramidal cells in the rat. Each CA3 cell receives approximately 46 mossy fibre inputs, so that the sparseness of this connectivity is thus 0.005%. By contrast, there are many more – possibly weaker – direct perforant path inputs also from layer 2 of the entorhinal cortex onto each CA3 cell, in the rat of the order of 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) (see Fig. 2). It is remarkable that the recurrent collaterals are distributed to other CA3 cells largely throughout the hippocampus [4,5,8,59,181], so that effectively the CA3 system provides a single network, with a connectivity of approximately 2% between the different CA3 neurons given that the connections are bilateral. The CA3–CA3 recurrent collateral system is even more extensive in macaques than in rats [72]. The neurons that comprise CA3, in turn, project to CA1 neurons via the Schaffer collaterals. In addition, projections that terminate in the CA1 region originate in layer 3 of the entorhinal cortex (see Fig. 1).

3.2. CA3 as an autoassociation or attractor memory

3.2.1. Arbitrary associations, and pattern completion in recall

Many of the synapses in the hippocampus show associative modification as shown by long-term potentiation, and this synaptic modification appears to be involved in learning see [11,83,99,100,101,107,108,177]. On the basis of the evidence summarized above, Rolls [118–121,123,124,126] and others [81,91,93] have suggested that the CA3 stage acts as an autoassociation memory which enables episodic memories to be formed and stored in the CA3 network, and that subsequently the extensive recurrent collateral connectivity allows for the retrieval of a whole representation to be initiated by the activation of some small part of the same representation (the cue). The crucial synaptic modification for this is in the recurrent collateral synapses. (A description of the operation of autoassociative networks is provided in detail elsewhere [10,55,133,136,148,149] including Memory, Attention, and Decision-Making (Rolls [145]))

The architecture of an autoassociation network is shown in Fig. 3, and the learning rule for the change in the synaptic weight is as shown in Eq. (1) [133,136,145].

\[ \delta w_{ij} = k \cdot r_i \cdot r_j' \] (1)

where \(k\) is a constant, \(r_i\) is the activation of the dendrite (the postsynaptic term), \(r_j'\) is the presynaptic firing rate, and \(w_{ij}\) is the synaptic weight.

The hypothesis is that because the CA3 operates effectively as a single network, it can allow arbitrary associations between inputs originating from very different parts of the cerebral cortex to be formed. These might involve associations between information originating in the temporal visual cortex about the presence of an object, and information originating in the parietal cortex.
about where it is. I note that although there is some spatial gradient in the CA3 recurrent connections, so that the connectivity is not fully uniform [59,181], 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, and the ability of attractor networks to operate with diluted connectivity shown in our computational studies prompted by this issue [168,169].

Crucial issues include how many memories could be stored in this system (to determine whether the autoassociation hypothesis leads to a realistic estimate of the number of memories that the hippocampus could store); whether the whole of a memory could be completed from any part; whether the autoassociation memory can act as a short-term memory, for which the architecture is inherently suited; and whether the system could operate with spatial representations, which are essentially continuous because of the continuous nature of space. These and related issues are considered in the remainder of Section 3.2 and in more detail elsewhere [141,145].

3.2.2. Storage capacity

We have performed quantitative analyses of the storage and retrieval processes in the CA3 network [169,170]. We have extended previous formal models of autoassociative memory [see 10] by analyzing a network with graded response units, so as to represent more realistically the continuously variable rates at which neurons fire, and with incomplete connectivity [132,168,169]. We have found, for example, that in general the maximum number \( p_{\text{max}} \) of firing patterns that can be (individually) retrieved is proportional to the number \( C_{\text{RC}} \) of (associatively) modifiable recurrent collateral synapses on to each neuron, by a factor that increases roughly with the inverse of the sparseness \( a \) of the neuronal representation. Each memory is precisely defined in the theory: it is a set of firing rates of the population of neurons (which represent a memory) that can be stored and later retrieved, with retrieval being possible from a fraction of the originally stored set of neuronal firing rates. The neuronal population sparseness \( a \) of the representation can be measured by extending the binary notion of the proportion of neurons that are firing to any one stimulus or event as

\[
a = \frac{1}{N} \sum \left( \frac{r_i}{N} \right)^2
\]

where \( r_i \) is the firing rate of the \( i \)th neuron in the set of \( N \) neurons. The sparseness ranges from \( 1/N \), when only one of the neurons responds to a particular stimulus (a local or grandmother cell representation), to a value of 1.0, attained when all the neurons are responding to a given stimulus. Approximately,

\[
p_{\text{max}} \approx \frac{C_{\text{RC}}}{a \ln(1/a)} k
\]

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 [169]. For example, for \( C_{\text{RC}} = 12,000 \) and \( a = 0.02 \), \( p_{\text{max}} \) is calculated to be approximately 36,000. This analysis emphasizes the utility of having a sparse representation in the hippocampus, for this enables many different memories to be stored. [The sparseness \( a \) in this equation is strictly the population sparseness \( a^p \). The population sparseness \( a^p \) would be measured by measuring the distribution of firing rates of all neurons to a single stimulus at a single time. The single neuron sparseness or selectivity \( a^s \) would be measured by the distribution of firing rates to a set of stimuli, which would take a long time. The selectivity or sparseness \( a^s \) of a single neuron measured across a set of stimuli often takes a similar value to the population sparseness \( a^p \) in the brain, and does so if the tuning profiles of the neurons to the set of stimuli are uncorrelated [38]. These concepts are elucidated by Franco, Rolls et al. (2007) [38].] (I note that the sparseness estimates obtained by measuring early gene changes, which are effectively population sparsenesses, would be expected to depend greatly on the range of environments or stimuli in which these were measured. If the environment was restricted to one stimulus, this would reflect the population sparseness. If the environment was changing, the measure from early gene changes would be rather undefined, as all the populations of neurons activated in an undefined number of testing situations would be likely to be activated.)

In order for most associative networks to store information efficiently, heterosynaptic long-term depression (as well as LTP) is required [35,125,133,136,145,169]. Simulations that are fully consistent with the analytic theory are provided by Simmen et al. [153] and Rolls et al. [132].

A number of points that arise, including measurement of the total amount of information (in bits per synapse) that can be retrieved from the network, the computational definition of a memory, the computational sense in which CA3 is an attractor network, and the possible computational utility of memory reconsolidation, are treated elsewhere [141,145]. Here I note that given that the memory capacity of the hippocampal CA3 system is limited, it is necessary to have some form of forgetting in this store, or other mechanism to ensure that its capacity is not exceeded. (Exceeding the capacity can lead to a loss of much of the information retrievable from the network.) Heterosynaptic LTD could help this forgetting by enabling new memories to overwrite old memories [130,145]. The limited capacity of the CA3 system does also provide one of the arguments that some transfer of information from the hippocampus to neocortical memory stores may be useful [see 171].

Given its limited capacity, the hippocampus might be a useful store for only a limited period, which might be in the order of days, weeks, or months. This period may well depend on the acquisition rate of new episodic memories. If the animal were in a constant and limited environment, then as new information is not being added to the hippocampus, the representations in the hippocampus would remain stable and persistent. These hypotheses have clear experimental implications, both for recordings from single neurons and for the gradient of retrograde amnesia, both of which might be expected to depend on whether the environment is stable or frequently changing. They show that the conditions under which a gradient of retrograde amnesia might be demonstrable would be when large numbers of new memories are being acquired, not when only a few memories (few in the case of the hippocampus being less than a few hundred) are being learned.

3.2.3. Recall and completion

A fundamental property of the autoassociation model of the CA3 recurrent collateral network is that the recall can be symmetric, that is, the whole of the memory can be retrieved and completed from any part [133,141,145]. For example, in an object–place autoassociative memory, an object could be recalled from a place retrieval cue, and vice versa. In a test of this, Day, Langston and Morris [27] trained rats in a study phase to learn in one trial an association between two flavors of food and two spatial locations. During a recall test phase they were presented with a flavor which served as a cue for the selection of the correct location. They found that injections of an NMDA receptor blocker (AP5) or AMPA/kainate receptor blocker (CNQX) to the dorsal hippocampus prior to the study phase impaired encoding, but injections of AP5 prior to the test phase did not impair the place recall, whereas injections of CNQX did impair the place recall. The interpretation is that somewhere in the hippocampus NMDA receptors are necessary for forming one-trial odor–place associations, and that recall can be performed without further involvement of NMDA receptors.
Evidence that the CA3 system is not necessarily required during recall in a reference memory spatial task, such as the water maze spatial navigation for a single spatial location task, is that CA3 lesioned rats are not impaired during recall of a previously learned water maze task [16,37]. However, if completion from an incomplete cue is needed, then CA3 NMDA receptors are necessary (presumably to ensure satisfactory CA3–CA3 learning) even in incomplete cue is needed, then CA3 NMDA receptors are necessary (presumably to ensure satisfactory CA3–CA3 learning) even in a reference memory task [49,106]. Thus, the CA3 system appears to be especially needed in rapid, one-trial object–place recall, and when completion from an incomplete cue is required (see further Section 5).

3.2.4. Continuous, spatial, patterns and CA3 representations

The fact that spatial patterns, which imply continuous representations of space, are represented in the hippocampus has led to the application of continuous attractor models to help understand hippocampal function. This has been necessary, because space is inherently continuous, because the firing of place and spatial view cells is approximately Gaussian as a function of the distance away from the preferred spatial location, because these cells have spatially overlapping fields, and because the theory is that these cells in CA3 are connected by Hebb-modifiable synapses. This specification would inherently lead the system to operate as a continuous attractor network. Continuous attractor network models have been studied by Amari [9], Zhang [183], Taylor [166], Samsonovich and McNaughton [151], Battaglia and Treves [13], Stringer et al. [159], Stringer et al. [158], Stringer et al. [160], Stringer and Rolls [157] and Rolls and Stringer [138] (see [136,145]), and are described briefly next.

A “Continuous Attractor” neural network (CANN) can maintain the firing of its neurons to represent any location along a continuous physical dimension such as spatial view, spatial position, head direction, etc. It uses excitatory recurrent collateral connections between the neurons (as are present in CA3) to reflect the distance between the neurons in the state space of the animal (e.g. place or head direction). These networks can maintain the bubble or packet of neural activity constant for long periods wherever it is started to represent the current state (head direction, position, etc.) of the animal, and are likely to be involved in many aspects of spatial processing and memory, including spatial vision. Global inhibition is used to keep the number of neurons in a bubble or packet of actively firing neurons relatively constant, and to help to ensure that there is only one activity packet.

Continuous attractor networks can be thought of as very similar to autoassociation or discrete attractor networks (see Rolls [145]), and have the same architecture, as illustrated in Fig. 3. The main difference is that the patterns stored in a CANN are continuous patterns, with each neuron having broadly tuned firing which decreases with for example a Gaussian function as the distance from the optimal firing location of the cell is varied, and with different neurons having tuning that overlaps throughout the space. Such tuning is illustrated in Fig. 4. For comparison, autoassociation networks normally have discrete (separate) patterns (each pattern implemented by the firing of a particular subset of the neurons), with no continuous distribution of the patterns throughout the space (see Fig. 4). A consequent difference is that the CANN can maintain its firing at any location in the trained continuous space, whereas a discrete attractor or autoassociation network moves its population of active neurons towards one of the previously learned attractor states, and thus implements the recall of a particular previously learned pattern from an incomplete or noisy (distorted) version of one of the previously learned patterns.

Space is continuous, and object representations are discrete. If these representations are to be combined in for example an object–place memory, then we need to understand the operation of networks that combine these representations. Rolls et al. [137] have shown that attractor networks can store both continuous patterns and discrete patterns (as illustrated in Fig. 4), and can thus be used to store for example the location in (continuous, physical) space (e.g. the place “out there” in a room represented by spatial view cells) where an object (a discrete item) is present. We showed this by storing associated continuous and discrete representations in the same single attractor network, and then showing that the representation in the continuous space could be retrieved by the discrete object that was associated with that spatial position; and that the representation of the discrete object could be retrieved by providing the position in the continuous representation of space.

If spatial representations are stored in the hippocampus, the important issue arises in terms of understanding memories that include a spatial component or context of how many such spatial representations could be stored in a continuous attractor network. The very interesting result is that because there are in general low correlations between the representations of places in different maps or charts (where each map or chart might be of one room or locale), very many different maps can be simultaneously stored in a continuous attractor network [13].

We have considered how spatial representations could be stored in continuous attractor networks, and how the activity can be maintained at any location in the state space in a form of short-term memory when the external (e.g. visual) input is removed. However, many networks with spatial representations in the brain can be updated by internal, self-motion (i.e. idiothetic), cues even when there is no external (e.g. visual) input. The ways in which path integration could be implemented in recurrent networks such as the CA3 system in the hippocampus or in related systems are described elsewhere [96,151,158,159], and have been applied to primate spatial view cells by Rolls and coworkers [138,160,161]. Cognitive maps [109] can be understood by the operations of these attractor

![Fig. 4.](image-url)
networks, and how they are updated by learning and by self-motion (Rolls [145]).

3.2.5. Mossy fibre inputs to the CA3 cells

We hypothesize that the mossy fibre inputs force efficient information storage by virtue of their strong and sparse influence on the CA3 cell firing rates [118–120,170]. (The strong effects likely to be mediated by the mossy fibres were also emphasized by McNaughton and Morris [91] and McNaughton and Nadel [92].) We (Rolls and Treves) [118–120,133,141,145,170] hypothesize that the mossy fibre input appears to be particularly appropriate in several ways. First, the fact that mossy fibre synapses are large and located very close to the soma makes them relatively powerful in activating the postsynaptic cell. Second, the firing activity of dentate granule cells appears to be very sparse [64,79] and this, together with the small number of connections on each CA3 cell, produces a sparse signal, which can then be transformed into sparse firing activity in CA3 by a threshold effect. The hypothesis is that the mossy fibre sparse connectivity solution performs the appropriate function to enable learning to operate correctly in CA3 [23,170]. The perforant path input would, the quantitative analysis shows, not produce a pattern of firing in CA3 that contains sufficient information for learning [170].

The particular property of the small number of mossy fibre connections onto a CA3 cell, approximately 46 (see Fig. 2), is that this has a randomizing effect on the representations set up in CA3, so that they are as different as possible from each other [119,120,133,141,145,170]. (This means for example that place cells in a given environment are well separated to cover the whole space.) The result is that any one event or episode will set up a representation that is very different from other events or episodes, because the set of CA3 neurons activated for each event is random. This is then the optimal situation for the CA3 recurrent collateral effect to operate, for it can then associate together the random set of neurons that are active for a particular event (for example an object in a particular place), and later recall the whole set from any part. It is because the representations in CA3 are unstructured, or random, in this way that large numbers of memories can be stored in the CA3 autoassociation system, and that interference between the different memories is kept as low as possible, in that they are maximally different from each other [58,133,169].

The requirement for a small number of mossy fibre connections onto each CA3 neuron applies not only to discrete [170] but also to spatial representations, and some learning in these connections, whether associative or not, can help to select out the small number of mossy fibres that may be active at any one time to select a set of random neurons in the CA3 [23]. Any learning may help by reducing the accuracy required for a particular number of mossy fibre connections to be specified genetically onto each CA3 neuron. The optimal number of mossy fibres for the best information transfer from dentate granule cells to CA3 cells is in the order of 35–50 [23,170]. The mossy fibres also make connections useful for feedforward inhibition [1].

On the basis of these and other points, we predicted that the mossy fibres may be necessary for new learning in the hippocampus, but may not be necessary for the recall of existing memories from the hippocampus [133,145,170]. Experimental evidence consistent with this prediction about the role of the mossy fibres in learning has been found in rats with disruption of the dentate granule cells [74] (Section 5).

We [141] have hypothesized that nonassociative plasticity of mossy fibres [see 14, 15] might have a useful effect in enhancing the signal-to-noise ratio, in that a consistently firing mossy fibre would produce nonlinearly amplified currents in the postsynaptic cell, which would not happen with an occasionally firing fibre [170]. This plasticity, and also learning in the dentate, would also have the effect that similar fragments of each episode (e.g. the same environmental location) recurring on subsequent occasions would be more likely to activate the same population of CA3 cells, which would have potential advantages in terms of economy of use of the CA3 cells in different memories, and in making some link between different episodic memories with a common feature, such as the same location in space. Consistent with this, dentate neurons that fire repeatedly are more effective in activating CA3 neurons [54].

As acetylcholine turns down the efficacy of the recurrent collateral synapses between CA3 neurons [48,53], then cholinergic activation also might help to allow external inputs rather than the internal recurrent collateral inputs to dominate the firing of the CA3 neurons during learning, as the current theory proposes. If cholinergic activation at the same time facilitated LTP in the recurrent collaterals (as it appears to in the neocortex), then cholinergic activation could have a useful double role in facilitating new learning at times of behavioral activation, when presumably it may be particularly relevant to allocate some of the limited memory capacity to new memories.

3.2.6. Perforant path inputs to CA3 cells

By calculating the amount of information that would end up being carried by a CA3 firing pattern produced solely by the perforant path input and by the effect of the recurrent connections, we have been able to show [170] that an input of the perforant path type, alone, is unable to direct efficient information storage. Such an input is too weak, it turns out, to drive the firing of the cells, as the “dynamics” of the network is dominated by the randomizing effect of the recurrent collaterals. On the other hand, an autoassociative memory network needs afferent inputs to apply the retrieval cue to the network. We have shown [170] that the perforant path system is likely to be the one involved in relaying the cues that initiate retrieval in CA3. The concept is that to initiate retrieval, a numerically large input is useful so that even a partial cue is sufficient; and that the retrieval cue need not be very strong, as the recurrent collaterals then take over in the retrieval process [145,170]. In contrast, during storage, strong signals, in the order of mV for each synaptic connection, are provided by the mossy fibre inputs to dominate the recurrent collateral activations, so that the new pattern of CA3 cell firing can be stored in the CA3 recurrent collateral connections [145,170].

3.3. Dentate granule cells

The theory is that the dentate granule cell stage of hippocampal processing which preceeds the CA3 stage acts as a competitive network in a number of ways to produce during learning the sparse yet efficient (i.e. non-redundant) representation in CA3 neurons that is required for the autoassociation implemented by CA3 to perform well [119,120,133,141,143,170]. An important property for episodic memory is that the dentate by acting in this way would perform pattern separation (or orthogonalization) [120,141,143,170], enabling the hippocampus to store different memories of even similar events, and this prediction has been confirmed [46,50,80,89,141,145] (Section 5).

As just described, the dentate granule cells could be important in helping to build and prepare spatial representations for the CA3 network. The actual representation of space in the primate hippocampus includes a representation of spatial view [144], whereas in the rat hippocampus it is of the place where the rat is. The representation in the rat may be related to the fact that with a much less developed visual system than the primate, the rat’s representation of space may be defined more by the olfactory and tactile as well as distant visual cues present, and may thus tend to reflect the place where the rat is. However, the spatial representations in the rat and primate could arise from essentially the same computational pro-
cess as follows [28,135]. The starting assumption is that in both the rat and the primate, the dentate granule cells (and the CA3 and CA1 pyramidal cells) respond to combinations of the inputs received. In the case of the primate, a combination of visual features in the environment will, because of the fovea providing high spatial resolution over a typical viewing angle of perhaps 10–20°, result in the formation of a spatial view cell, the effective trigger for which will thus be a combination of visual features within a relatively small part of space. In contrast, in the rat, given the very extensive visual field subtended by the retinal retina, which may extend over 180–270°, a combination of visual features formed over such a wide visual angle would effectively define a position in space that is a place [28].

The entorhinal cortex contains grid cells, which have high firing in the rat in a two-dimensional spatial grid as a rat traverses an environment, with larger grid spacings in the ventral entorhinal cortex [39,51]. This may be a system optimized for path integration [96] which may self-organize during locomotion with longer time constants producing more widely spaced grids in the ventral entorhinal cortex [73]. How are the grid cell representations, which would not be suitable for association of an object or reward with a place to form an episodic memory, transformed into a place representation that would be appropriate for this type of episodic memory? I have proposed that this could be implemented by a competitive network [145] in the dentate gyrus which operates to form place cells, implemented by each dentate granule cell learning to respond to particular combinations of entorhinal cortex cells firing, where each combination effectively specifies a place, and this has been shown to be feasible computationally [143]. The sparse representations in the dentate gyrus, implemented by the mutual inhibition through inhibitory interneurons and competitive learning, help to implement this ‘pattern separation’ effect [119,120,133,145].

In primates, spatial view cells represent a scene view allocentrically, as described in Section 4. How could such spatial view representations be formed, in which the relative spatial position of features in a scene is encoded? I have proposed that this involves competitive learning analogous to that used to form place cells in rats, but in primates operating on the representations of objects that reach the hippocampus from the inferior temporal visual cortex [146]. We have shown that in complex natural scenes the receptive fields of inferior temporal cortex neurons become reduced in size and asymmetric with respect to the fovea [2,147], and Rolls et al. [146] have demonstrated in a unifying computational approach that competitive network processes operating in areas such as the parahippocampal cortex, the entorhinal cortex, and/or the dentate granule cells could form unique views of scenes by forming a sparse representation of these object or feature tuned inferior temporal cortex ventral visual stream representations which have some spatial asymmetry.

3.4. CA1 cells

The CA3 cells connect to the CA1 cells by the Schaeffer collateral synapses. The associative modifiability in this connection helps the full information present in CA3 to be retrieved in the CA1 neurons [128,152,171,172]. Part of the hypothesis is that the separate sub-parts of an episodic memory, which must be represented separately in CA3 to allow for completion, can be combined together by competitive learning in CA1 to produce an efficient retrieval cue for the recall via the backprojection pathways to the neocortex of memories stored in the neocortex [120,121,128,129,171].

3.5. Backprojections to the neocortex, and memory recall

The need for information to be retrieved from the hippocampus to affect other brain areas was noted in Section 1. The way in which this could be implemented via backprojections to the neocortex is now considered.

It is suggested that the modifiable connections from the CA3 neurons to the CA1 neurons allow the whole episode in CA3 to be produced in CA1. The CA1 neurons would then activate, via their termination in the deep layers of the entorhinal cortex, at least the pyramidal cells in the deep layers of the entorhinal cortex (see Fig. 1). These entorhinal cortex layer 5 neurons would then, by virtue of their backprojections [75,179] to the parts of cerebral cortex that originally provided the inputs to the hippocampus, terminate in the superficial layers (including layer 1) of those neocortical areas, where synapses would be made onto the distal parts of the dendrites of the (superficial and deep) cortical pyramidal cells [119–121]. The areas of cerebral neocortex in which this recall would be produced could include multimodal cortical areas (e.g. the cortex in the superior temporal sulcus which receives inputs from temporal, parietal and occipital cortical areas, and from which it is thought that cortical areas such as 39 and 40 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, which by building new long-term and structured representations can be considered as a form of memory consolidation [119–121,123,124,145], or in organizing cell actions.

The hypothesis 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 I suggest that the backprojection synapses are set up to have the appropriate strengths for recall is as follows [119–121]. During the setting up of a new episodic memory, there would be strong feedforward activity progressing towards the hippocampus. During the episode, the CA3 synapses would be modified, and via the CA1 neurons and the subiculum, a pattern of activity would be produced on the backprojecting synapses to the entorhinal cortex. Here the backprojecting synapses from active backprojection axons onto pyramidal cells being activated by the forward inputs to entorhinal cortex would be associatively modified. A similar process would be implemented at preceding stages of neocortex, that is in the parahippocampal gyrus/perirhinal cortex stage, and in association cortical areas, as shown in Fig. 1.

The concept is that, during the learning of an episodic memory, cortical pyramidal cells in at least one of the stages would be driven by forward inputs, but would simultaneously be receiving backprojected activity (indirectly) from the hippocampus which would by pattern association from the backprojecting synapses to the cortical pyramidal cells become associated with whichever cortical cells were being made to fire by the forward inputs. Then later on, during recall, a recall cue from perhaps another part of cortex might reach CA3, where the firing during the original episode would be completed. The resulting backprojecting activity would then, as a result of the pattern association learned previously, bring back the firing in any cortical area that was present during the original episode. Thus retrieval involves reinstating the activity that was present in different cortical areas that was present during the learning of an episode. (The pattern association is also called heteroassociation to contrast it with autoassociation. The pattern association operates at multiple stages in the backprojection pathway, as made evident in Fig. 1.) If the recall cue was an object, this might result in recall of the neocortical firing that represented the place in which that object had been seen previously. As noted elsewhere in this paper
and by McClelland et al. [88], that recall might be useful to the neocortex to help it build new semantic memories, which might inherently be a slow process and is not part of the theory of recall.

A plausible requirement for a successful hippocampal-directed recall operation is that the signal generated from the hippocampally retrieved pattern of activity, and carried backwards towards neocortex, remain undegraded when compared to the noise due, at each stage, to the interference effects caused by the concurrent storage of other patterns of activity on the same backprojecting synaptic systems. That requirement is equivalent to that used in deriving the storage capacity of such a series of heteroassociative memories, and it was shown by Treves and Rolls [169,171] that the maximum number of independently generated activity patterns that can be retrieved is given, essentially, by the same formula as (3) above where, however, \( a \) is now the sparseness of the representation at any given stage, and \( C \) is the average number of (back-)projections each cell of that stage receives from cells of the previous one. \( (k' \) is a similar slowly varying factor to that introduced above.) If \( p \) is equal to the number of memories held in the hippocampal memory, it is limited by the retrieval capacity of the CA3 network, \( p_{max} \). Putting together the formula for the latter with that shown here, one concludes that, roughly, the requirement implies that the number of afferents of (indirect) hippocampal origin to a given neocortical stage (\( \text{HBP} \)) must be \( \text{HBP} = \frac{C\max}{C_{max}/a_{CA3}} \), where \( C \) is the number of recurrent collaterals to any given cell in CA3, the average sparseness of a representation is \( \alpha_{CA3} \). The required divergence can be kept within reasonable limits only by assuming that the backprojecting system of CA3 cells). The required divergence can be kept within reasonable limits only by assuming that the backprojecting system provides a quantitative account of why the cerebral cortex has as many backprojections as forward projection connections. Further aspects of the operation of the backprojecting systems are described elsewhere [145].

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, \( \text{HBP} \) should be as large as \( \frac{C\max}{C_{max}/a_{CA3}} \), i.e. 12,000 in the rat. If then \( \text{HBP} \) has to be of the same order as \( C \), one is led to a very definite conclusion: a mechanism of the type envisaged here could not possibly rely on a set of monosynaptic CA3-to-neocortex backprojections. This would imply that, to make a sufficient number of synapses on each of the vast number of neocortical cells, each cell in CA3 has to generate a disproportionate number of synapses (i.e. \( \text{HBP} \) times the ratio between the number of neocortical and that of CA3 cells). The required divergence can be kept within reasonable limits only by assuming that the backprojecting system is polysynaptic, provided that the number of cells involved grows gradually at each stage, from CA3 back to neocortical association areas [171] (cf. Fig. 1).

The theory of recall by the backprojections thus provides a new approach to hippocampal function I am taking here is that based on phenomena such as theta and gamma oscillations that has been investigated by Lisman and Redish [82].

3.6. Temporal order memory in the hippocampus, and episodic memory

There has for some time been evidence that the hippocampus plays a role in temporal order memory, even when there is no spatial component [56,67,141]. In humans, the hippocampus becomes activated when the temporal order of events is being processed [78]. If this is a function that can be implemented in the hippocampus, I now propose that this could be very important for understanding episodic memory, which often may comprise a temporal sequence of events. However, until the last few weeks it has not been clear to me how temporal order memory could be implemented in the hippocampus, or, for that matter, in other brain structures. However, I have now developed the outline of a theory of temporal order memory, and how it could be implemented in the hippocampus [149], as follows.

The approach is based on recent neurophysiological evidence of MacDonald and Eichenbaum [84] showing that neurons in the rat hippocampus have firing rates that reflect which temporal part of the task is current. In particular, a sequence of different neurons is activated at successive times during a time delay period (see Rolls and Deco [149]). The tasks used included an object–odor paired associate non-spatial task with a 10 s delay period between the visual stimulus and the odor. The new evidence also shows that a large proportion of hippocampal neurons fire in relation to individual events in a sequence being remembered (e.g. a visual object or odor), and some to combinations of the event and the time in the delay period [84].

These interesting neurophysiological findings indicate that rate encoding is being used to encode time, that is, the firing rates of different neurons are high at different times within a trial, delay period, etc. [84] (see Rolls and Deco [149]). This provides the foundation for a new computational theory of temporal order memory within the hippocampus (and also the prefrontal cortex) which I call temporal order memory and which utilizes the slow transitions from one attractor to another which are a feature that arises at least in some networks in the brain due to the noise-influenced transitions from one state to another.

First, because some neurons fire at different times in a trial, the time in which an object (e.g. a visual stimulus or odor) was presented could become encoded in the hippocampus by an association implemented in the CA3 recurrent collaterals between the neurons that represent the object (already known to be present in the hippocampus for tasks for which the hippocampus is required [140,144]) and the ‘time encoding’ neurons in the hippocampus [84]. This would allow associations for the time at which the object was present to be formed.

Second, these associations would provide the basis for the recall of the object from the time in a trial, or vice versa. The retrieval of object or temporal information from each other would occur in a way that is analogous to that shown for recalling the object from the place, or the place from the object [137], but substituting the details of the properties of the ‘time encoding’ neurons for what was previously the spatial (place) component. In addition, if the time encoding neurons simply cycled through their normal sequence during recall, this would enable the sequence of objects or events associated with each subset of time encoding neurons to be recalled correctly in the order in which they were presented.

Third, we need a theory of what the origin of the temporal effect whereby different hippocampal (or potentially prefrontal cortex) neurons fire in different parts of a trial or delay period. The properties of the ‘time encoding neurons’ [84] (see Rolls and Deco [149]) are key here, and we need to understand how they are generated. Are they generated within the hippocampus, or elsewhere, and in any case, what is the mechanism by which different neurons have high firing rates at different times in a trial? The fundamentally new approach to hippocampal function I am taking here is that rate encoding is being used, that is, the firing rates of different neurons are high at different times within a trial [84] (see Rolls and Deco [149]). This is a radically different approach to order encoding than that based on phenomena such as theta and gamma oscillations that has been investigated by Lisman and Redish [82].

We can consider three hypotheses about how the firing of the ‘time encoding’ hippocampal neurons is produced. All utilize slow transitions between attractor states that can be a property of noisy attractor networks. The first hypothesis is that an attractor network with realistic dynamics (modelled at the integrate-and-fire level with a dynamical implementation of the neuronal membrane and synaptic current dynamics, and with synaptic or neuronal adaptation) can implement a sequence memory, as shown by Deco and Rolls [31]. The hypothesis is that there are several different
attractors, and that there are weak connections between the different attractors. In the model, adaptation produces effects whereby whatever sequence (order of stimuli) is presented on an individual trial, that order can be replayed in the same sequence because as one attractor state dies as a result of the adaptation, the next attractor to emerge from the spontaneous firing because of the spiking-related noise is the one that has been active least recently, as it is the one that is least adapted [31]. The whole system operates at a rather slow timescale for the transitions between the attractors partly because of the time for the noise to drive the system from one attractor state to another, and the slow time course of the adaptation [31,149]. This implements a type of order memory.

The second hypothesis is analogous, and is also implemented in a recurrently connected system such as the hippocampal CA3 system or local recurrent circuits in the neocortex [149]. This second theory is that again there are several attractors, but that each attractor is connected by slightly stronger forward than reverse synaptic weights to the next. In previous work, we have shown that with an integrate-and-fire implementation with spiking noise this allows slow transitions from one attractor state to the next [29,30]. During learning of the synaptic weights in the network, adaptation might lead to each ‘time encoding’ population of neurons responding for only a limited period, helping to produce multiple sequentially activated populations of time encoding neurons [84] (see Rolls and Deco [149]). In this scenario, an associative pool of neurons is unlikely to be helpful, and stronger forward than reverse weights between different attractors each consisting of a different population of ‘time encoding’ neurons would be the essence. It will be of interest to investigate whether this system, because of the noise, is limited to transitions between up to perhaps $7 \pm 2$ different sequential firing rate states with different neuronal subpopulations for each state, and thus provides an account for the limit of the numerical number $7 \pm 2$ on short-term memory and related types of processing [97], and for the recency part of short-term memory in which the items are naturally recalled in the order in which they were presented. This is the most likely model at present of short-term memory and its natural propensity to store and to recall items in the order in which they were received [149].

A variation on this implementation that I have proposed would be to have short-term attractor memories with different time constants (for example of adaptation), but all started at the same time [149]. This could result in some attractors starting early in the sequence and finishing early, and with other attractors starting up a little later, but lasting for much longer in time. The neurons recorded by MacDonald and Eichenbaum [84] (see Rolls and Deco [149]) are not inconsistent with this possibility. This type of time encoding representation could also be used to associate with items to implement an item–order memory.

It is thus suggested that temporal order memory could be implemented in the hippocampus in this way and could make an important contribution to episodic memory in which several events linked in the correct order might form an episode. The theory shows how items in a particular temporal order could be separated from each other, a property we have referred to as the temporal pattern separation effect [141]. The theory of episodic memory described here shows how events and sequences of events could be recalled from the hippocampus to the neocortex, and there a longer term more semantic representation of a recalled episode, such as what happened on one’s fifth birthday, might be stored, and then accessed to describe the episode. For the order to be correctly implemented in the semantic neocortical store, a similar mechanism, involving for example stronger forward than reverse synaptic weights between long-term memory representations in attractors, could build an appropriate long-term order memory [149].

4. Systems-level neurophysiology of the primate hippocampus

The systems-level neurophysiology of the hippocampus shows what information could be stored or processed by the hippocampus. To understand how the hippocampus works it is not sufficient to state just that it can store information – one needs to know what information. The systems-level neurophysiology of the primate hippocampus has been reviewed by Rolls and Xiang [144], and a summary is provided here because it provides a perspective relevant to understanding the function of the human hippocampus that is somewhat different from that provided by the properties of place cells in rodents, which have been reviewed elsewhere [62,63,90,103,110].

4.1. Spatial view neurons in the primate hippocampus

We have shown that the primate hippocampus contains spatial cells that respond when the monkey looks at a certain part of space, for example at one quadrant of a video monitor while the monkey is performing an object–place memory task in which he must remember where on the monitor he has seen particular images [122,135]. Approximately 9% of the hippocampal neurons have such spatial view fields, and about 2.4% combine information about the position in space with information about the object that is in that position in space [122]. The representation of space is for the majority of hippocampal neurons not egocentric but allocentric coordinates [36]. These spatial view cells can be recorded while monkeys move themselves round the test environment by walking (or running) on all fours [45,117,131,134]. These hippocampal ‘spatial view neurons’ respond significantly differently for different allocentric spatial views and have information about spatial view in their firing rate, but do not respond differently just on the basis of eye position, head direction, or place. If the view details are obscured by curtains and darkness, then some spatial view neurons (especially those in CA1 and less those in CA3) continue to respond when the monkey looks towards the spatial view field, showing that these neurons can be updated for at least short periods by idiothetic (self-motion) cues including eye position and head direction signals [117,132].

4.2. Object–place neurons in the primate hippocampus

A fundamental question about the function of the primate including human hippocampus in relation to episodic memory is whether object as well as allocentric spatial information is represented. To investigate this, Rolls et al. [140] made recordings from single hippocampal formation neurons while macaques performed an object–place memory task that required the monkeys to learn associations between objects, and where they were shown in a room. Some neurons (10%) responded differently to different objects independently of location; other neurons (13%) responded to the spatial view independently of which object was present at the location; and some neurons (12%) responded to a combination of a particular object and the place where it was shown in the room. These results show that there are separate as well as combined representations of objects and their locations in space in the primate hippocampus. This is a property required in an episodic memory system, for which associations between objects and the places where they are seen are prototypical. The results thus show that a requirement for a human episodic memory system, separate and combined neuronal representations of objects and where they are seen “out there” in the environment, is present in the primate hippocampus [140].

What may be a corresponding finding in rats is that some rat hippocampal neurons respond on the basis of the conjunc-
There are backprojections from the hippocampus to the entorhinal cortex and thus to parahippocampal areas, and these backprojections could enable the hippocampus to influence the spatial representations found in the entorhinal cortex and parahippocampal gyrus. On the other hand, some of the spatial functions may be provided for these parahippocampal areas, which will in turn influence the hippocampus. However, it is argued below that the hippocampus may be able to make a special contribution to event or episodic memory, by enabling in the CA3 network with its very widespread recurrent collateral connections an association between any one item with any other item to form an arbitrary association to represent an event.

4.3. Recall-related neurons in the primate hippocampus

It has now been possible to investigate directly, neurophysiologically, the hippocampal recall process in primates [144]. We used a visual object–place memory task because this is prototypical of episodic memory. It has been shown that a one-trial odor–place recall memory task is hippocampal-dependent in rodents [27]. We designed a one-trial object–place recall task, in which the whole memory was recalled from a part of it. The task is illustrated in Fig. 5. Images of new objects were used each day, and within a day the same objects were used, so that with non-trial–unique objects within a day, the recall task is quite difficult.

Recordings were made from 347 neurons in the hippocampus of a macaque performing the object–place recall task. The following types of neurons were found in the task [144].

One type of neuron had responses that occurred to one of the objects used in the task. A number of these neurons had activity that was related to the recall process. An example of one of these neurons is shown in Fig. 6. The neuron had activity that was greater to object one not only when it was shown in Stages 1–3 of the task, but also in the delay period following Stage 3 when the object was no longer visible, and in Stage 4, when the object was no longer visible and the macaque was touching the remembered location of the object. Thus while the location was being recalled from the object, this type of neuron continued to respond as if the object was present, that is it kept the representation of the object active after the object was no longer visible, and the place to touch was being recalled. 16 of the neurons responded in this way, and an additional 6 had object-related firing that did not continue following Stage 3 of the task in the recall period. The difference of the firing rates of these 22 neurons to the different places was in many cases highly statistically significant (e.g. \( p < 10^{-6} \)) [144]. None of these neurons had differential responses for the different places used in the object–place recall task.

A second type of neuron had responses related to the place (left or right) in which an object was shown in Stage 1 or 2 of each trial. An example of one of these neurons is shown in Fig. 7. The neuron responded more when an object was shown in the left position (P1) than in the right position (P2) on the screen. Interestingly, when the recall object was shown in Stage 3 of the trial in the top centre of the screen, the neuron also responded as if the left position (P1) was being processed on trials on which the left position had to be recalled. This firing continued in the delay period after the recall cue had been removed at the end of Stage 3, and into Stage 4. Thus this type of neuron appeared to reflect the recall of the position on the screen at which the object had been represented. Analysis of trials on which errors were made indicated that the responses were not just motor response related, for if due to some response bias the monkey touched the incorrect side, the neuron could still respond according to the correct recalled location. 13 neurons had differential responses to the different places P1 and P2, and continued to show place-related activity in the recall part of the task, Stage 3.
consistent with these findings, the object-related neurons could be part of the completion of the object after it had disappeared in the recall phase of the task. By continuing to respond to the object while the place is being recalled in the task, the neurons with continuing activity to the object could reflect the operation of the object–place recall process that is hypothesized to take place in the CA3 cells. By continuing to respond to the object when a non-spatial recall stimulus was being shown, a memory recall component, in that they did not respond in Stage 3 when the object being shown as a place recall cue. The neuron responded more when place 1 was the correct place to be recalled on a trial. The fourth histogram pair shows the neuronal responses in Stage 4 when the macaque was recalling and touching the place at which the cue recall object had been shown. The responses of the neuron were place-related even in Stage 3 when the object being shown as a place recall cue was at the top of the screen, in the delay after Stage 3 of the task, and in Stage 4.

Five other neurons had left–right place-related responses without a memory recall component, in that they did not respond in Stage 3 of the task, when a non-spatial recall stimulus was being shown, and a place should be being recalled (see Table 1). The population of 18 neurons as a population had statistically significant place-related responses [144]. The new finding is that 13 of the neurons had place-related responses when a place was being recalled by an object cue.

The responses of the population of neurons recorded in one macaque are shown in Table 1. In addition to the neurons described above, 3 further neurons responded to particular combinations of objects and places, e.g. to object 1 when it was shown in place 1, but not to other combinations.

The recording sites of the object and of the place neurons were within the hippocampus proper [144]. The mean firing rate of the population of responsive neurons (see Table 1) to the most effective object or place was 7.2 ± 0.6 spikes/s (±SEM), and their mean spontaneous rate was 3.2 ± 0.6 spikes/s.

These findings [144] are the first we know in the primate hippocampus of neuronal activity that is related to recall. It is particularly interesting that the neurons with continuing activity to the object after it had disappeared in the recall phase of the task could reflect the operation of the object–place recall process that is hypothesized to take place in the CA3 cells. By continuing to respond to the object while the place is being recalled in the task, the object-related neurons could be part of the completion of the whole object–place combination memory from an autoassociation or attractor process in CA3 [141]. Consistent with these findings, and with the computational theory, it has now been reported that human hippocampal neurons are activated during recall [44].

The neurons with recall-related activity in the object–place recall task also provide neurophysiological evidence on the speed of association learning in the hippocampal formation. Given that this is a one-trial object–place recall task, with the association between the object and its place being made in Stages 1 and 2 of each trial (see Fig. 5), it is clear that it takes just one trial for the object–place associations to be formed that are relevant to the later recall on that trial. This is the speed of learning that is required for episodic memory, and this neurophysiological evidence shows that this type of rapid, one-trial object–place learning is represented in the primate hippocampus.

4.4. Reward–place neurons in the primate hippocampus

The primate anterior hippocampus (which corresponds to the rodent ventral hippocampus) receives inputs from brain regions involved in reward processing such as the amygdala and orbitofrontal cortex [114]. To investigate how this effective input may be incorporated into primate hippocampal function, Rolls and Xiang [139] recorded neuronal activity while macaques performed a reward–place association task in which each spatial scene shown on a video monitor had one location which if touched yielded a preferred fruit juice reward, and a second location which yielded a less preferred juice reward. Each scene had different locations for the different rewards. Of 312 hippocampal neurons analyzed, 18% responded more to the location of the preferred reward in different scenes, and 5% to the location of the less preferred reward [139]. When the locations of the preferred rewards in the scenes were reversed, 60% of 44 neurons tested reversed the location to which they responded, showing that the reward–place associations could be altered by new learning in a few trials. The majority (82%) of these 44 hippocampal reward–place neurons tested did not respond to object–reward associations in a visual discrimination object–reward association task. Thus the primate hippocampus contains a representation of the reward associations of places “out there” being viewed, and this is a way in which effective information can be stored as part of an episodic memory, and how the current mood state may influence the retrieval of episodic memories. There is consistent evidence that rewards available in a spatial environment can influence the responsiveness of rodent place neurons [57,165].

5. Tests of the theory

5.1. Dentate granule cells

The theory predicts that the dentate granule cell mossy fibre system of inputs to the CA3 neurons is necessary to store spatial memories, but not to recall them. Lassalle et al. [74] have obtained evidence consistent with this in rats with damage to the mossy fibre system, and there is further evidence consistent with this [26,77,141]. The theory predicts that pattern separation is performed by the dentate granule cells. Evidence consistent with this has been found neurophysiologically in the small sparsely encoded place fields of dentate neurons [64,80] and their reflection in CA3 neurons [80]. It has been shown that selective dentate lesions in rats [46,47,50,141,145] or dentate NMDA receptor knockouts in mice [89] impair spatial, object–place (or reward–place: remembering where to find a reward) association tasks especially when the places are close together and require pattern separation before storage in CA3.

**Table 1**

| Object with activity continuing after the recall cue | 16 |
| Object with activity not continuing after the recall cue | 6 |
| Place with activity during and after the recall cue | 13 |
| Place with activity during the recall cue | 3 |
| Object × place | 3 |
| Total | 347 |
If adult neurogenesis in the dentate gyrus does prove to be functionally relevant, its computational role could be to facilitate pattern separation for new patterns, by providing new dentate granule cells with new sets of random connections to CA3 neurons. Consistent with the dentate spatial pattern separation hypothesis [119,120,129,170,171], in mice with impaired dentate neurogenesis, spatial learning in a delayed non-matching-to-place task in the radial arm maze was impaired for arms that were presented with little separation, but no deficit was observed when the arms were presented farther apart [24]. Consistently, impaired neurogenesis in the dentate also produced a deficit for small spatial separations in an associative object-in-place task [24].

The theory predicts that the direct perforant path input from the entorhinal cortex to the CA3 cells (which bypasses the dentate granule cells) is involved in the recall of memory from the CA3 system, and Lee and Kesner [77] have obtained evidence consistent with this in a Hebb–Williams maze recall task.

5.2. CA3

The theory predicts that the CA3 is especially important in object–place or reward–place tasks in which associations must be formed between any spatial location and any object (referred to as arbitrary associations). There is much evidence from subregion analyses involving disruption of CA3 that CA3 is necessary for arbitrary associations between places and objects or rewards [47,141]. Similar impairments were obtained following deletion of CA3 NMDA receptors in mice in the acquisition of an odor–context paired associate learning task [115]. If place or time is not a component, associative tasks such as odor–object association are not impaired [141], underlining the fact that the hippocampus is especially involved in episodic types of associative memory which typically involve place and/or time.

The theory predicts that the CA3 is especially important in object–place or reward–place completion tasks, in which associations must be completed from a part of the whole. It has been shown that if completion from an incomplete cue is needed, then CA3 NMDA receptors are necessary (presumably to ensure satisfactory CA3–CA3 learning) even in a reference memory task [49,106].

The theory predicts that the CA3 system is especially needed in rapid, one-trial object–place learning and recall. It has been shown that hippocampal NMDA receptors (necessary for long-term potentiation to occur) are needed for one-trial flavor–place association learning, and that hippocampal AMPA/kainate receptors are sufficient for the recall, though the hippocampal subregion involved was not tested [27]. In subregion studies, Kesner et al. have shown that CA3 lesions produce chance performance on a one-trial object–place recall task [69] and other object–spatial tasks [66,141]. For example, CA3 lesions produce chance performance on both a one-trial object–place recall and place–object recall task [69]. This is evidence that CA3 supports arbitrary associations as well as episodic memory based on one-trial learning. A control fixed visual conditional to place task with the same delay was not impaired, showing that it is recall after one-trial (or rapid, episodic) learning that is impaired [69]. CA3 NMDA receptors are as predicted by the theory necessary for rapid/one-trial spatial learning, as shown by a mouse knockout study by Nakazawa and coworkers [107,108,167]. As described in Section 4, we have shown that hippocampal CA3 neurons reflect the computational processes necessary for one-trial object–place event memory, used as a model for episodic memory [144].

The theory predicts that if primates including humans can form an episodic memory in which objects or people are seen at particular locations even though the observer viewing the space and has never been to those locations ‘out there’ in space, there should be a neural system in CA3 that can support such associations between places ‘out there’ in a scene and objects. Exactly this is provided by the spatial view neurons Rolls and colleagues have discovered that are present in CA3 [45,117,131,134,139,140,144]. Place cells will not do for this type of episodic memory.

5.3. Recall via CA1 to neocortex

The theory shows quantitatively, analytically, how memories could be retrieved from the hippocampus to the neocortex [171], and this has been shown by simulation of the multistage hippocampal system including the entorhinal cortex, dentate, CA3, CA1, and return to the entorhinal cortex to recall the memory to be quantitatively realistic [128].

Many further tests of the theory are described elsewhere [70,141,145].

6. Discussion

The present theory holds that the hippocampus is used for the formation of episodic memories using autoassociation. This function is often necessary for successful spatial computation, but is not itself spatial computation. Instead, I believe that spatial computation is more likely to be performed in the neocortex (utilizing information if necessary for the particular task 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 and parahippocampal cortex can lead to problems, such as topographical and other spatial agnosias, in humans [see 71]. This is consistent with spatial computations normally being performed in the neocortex. In monkeys, there is evidence for a role of the parietal cortex in allocentric spatial computation. For example, monkeys with parietal cortex lesions are impaired at performing a landmark task, in which the object to be chosen is signified by the proximity to it of a ‘landmark’ (another object) [173].

The key aspect of the hippocampal theory, which relates it very closely to episodic memory, is that the hippocampal system has the dentate granule to mossy fibre to CA3 cell system, which with its sparse connectivity (46 synapses onto each CA3 neuron) has a randomizing effect on the representations set up in CA3, so that they are as different as possible from each other; and that this then enables each associative memory formed in the CA3 recurrent collaterals to be as different as possible from all the other event memories (see Section 3.3). This is then the optimal situation for the CA3 recurrent collateral effect to operate, for it can then associate together the random set of neurons that are active for a particular event (for example an object in a particular place), and later recall the whole set from any part. It is because the representations in CA3 are unstructured, or random, in this way that large numbers of memories can be stored in the CA3 autoassociation system, and that interference between the different memories is kept as low as possible, in that they are maximally different from each other. This is the essence of the hippocampus as a memory system suitable for unstructured memories such as event or episodic memories, each of which should be kept as distinct from the others as possible.

In this theory, the entorhinal cortex grid cell to dentate/CA3 place cell transformation that can be achieved by competitive learning [143] is necessary so that a sparse representation of a place with different representations for different places suitable for association with an object or reward can be made available for CA3. The grid cell representation does not have this property, in that the representations of different places by grid cells are too highly correlated to be useful in an associative memory [145]. The utility of the entorhinal grid cell system is likely to be its role in path integra-
tion [96]; in the idiothetic (self-motion) update of location, where the location is represented as a series of ring attractors of different spatial sizes, with no beginning and no end to each ring continuous attractor. Because a ring is endless, and the packet of activity just keeps going round the ring as locomotion continues, the system can perform path integration beyond any known environment, and in the dark, where the spatial environmental cues are not known or cannot be seen, so allowing for example return to base.

Once place cell representations have been set up in the hippocampus, they form a continuous spatial attractor network because of associative synaptic modification during locomotion producing synapses that are stronger the nearer together two places are, and thus the more likely coactivity between CA3 pyramidal cells is [145,158,160–162]. Although this is a continuous representation, as it has to be because space is continuous, it is suitable for association with the representations of objects, which are discrete, to form an episodic memory [137]. An important property of this spatial system is that there can be many separate maps, or charts, each of a different environment, stored in the hippocampus [13], as each chart has low correlations with any other chart, due in part to the randomizing effect of the mossy fibres. The essential property of the hippocampal CA3 system then may be that it is not specialized to perform the computations necessary for spatial navigation [18,94,95,111], but instead is a memory system that inevitably contains some continuous representations of space because places, landmarks etc. are being encoded as part of an episodic memory system, and space is inherently continuous. Having said this, the hippocampus according to the present (Rolls’) theory does form memories of where objects, people, rewards etc. have been found in space, and so may be useful when one is looking for a particular object or reward, for the place can be recalled from the hippocampal representation. Once recalled to the neocortex, the place may then be used in a neocortically organized system to enable actions to be performed to reach a place, which could involve spatial navigation, using landmark or body centred algorithms, but also actions such as getting on a boat or bus, taking a plane, etc.

The hippocampus is the only part of the brain I know with the type of sparse connectivity implemented in the mossy fibre to CA3 synaptic connections to set up randomly different representations. It sets up a fascinating comparison with the neocortex. In the neocortex, the connectivity is set up to emphasize feedforward connectivity from one area to the next that has large numbers of connections onto each neuron, in the order of several thousand. This enables the neocortex to form representations by competitive learning that represent combinations of the ‘features’ present at an earlier stage, and which are separated apart in the space. Such a process operating over several stages of a hierarchy enables representations of different objects to be formed. Here the representations are not arbitrarily and randomly different, as in CA3, but instead represent convergent of information over large information spaces (using many synapses onto each neuron) that together enables an object to be ‘diagnosed’, that is, represented, as in Rolls’ theory of invariant visual object recognition in the ventral cortical visual stream [127,142,145,147,176]. In the neocortical system the local recurrent collateral connections between nearby cortical pyramidal cells play a largely different role to that in the hippocampal theory. In the neocortex they implement short-term memory which provides the basis for planning and attention; and decision-making including categorization [145,149].

A theory closely related to the present theory of how the hippocampus operates has been developed by McClelland et al. [88]. It is very similar to the theory we have developed [118–121,145,170,171] at the systems level, except that it takes a stronger position on the gradient of retrograde amnesia, emphasizes that recall from the hippocampus of episodic information is used to help build semantic representations in the neocortex, and holds that the last set of synapses that are modified rapidly during the learning of each episode are those between the CA3 and the CA1 pyramidal cells, as described above (see Fig. 1). It also emphasizes the important point that the hippocampal and neocortical memory systems may be quite different, with the hippocampus specialized for the rapid learning of single events or episodes, and the neocortex for the slower learning of semantic representations which may necessarily benefit from the many exemplars needed to shape the semantic representation.

In this paper we have thus seen that there is neurophysiological evidence that different representations of the type important in episodic memory, including object and place, and reward and place, are brought together in the primate hippocampus, and even that the whole representation can be completed from a partial retrieval cue in a one-trial object–place memory task. It appears to be a property of the hippocampus that it is involved in associations when one of the associates is place or time [141,145,149]. We have moreover seen that the representation of space in the primate hippocampus provided by spatial view cells is appropriate for episodic memory in primates including humans, for it is a representation of space “out there”, which is prototypical of the spatial representation that is involved in episodic memory in primates including humans. We have also described and updated a theory of how the hippocampal system could implement episodic memory [118–121,124,126,128,129,133,141,145,149,169–171] and shown how there is now considerable empirical support for the theory [70,141,145].

Acknowledgements

Different parts of the research described here were supported by Programme Grants from the Medical Research Council, by a Human Frontier Science program grant, by an EEC BRAIN grant, by the MRC Oxford Interdisciplinary Research Centre in Cognitive Neuroscience, and by the Oxford McDonnell–Pew Centre in Cognitive Neuroscience. The author has performed the experimental and theoretical work which is incorporated in some of the ideas presented here on the hippocampus with many colleagues, including Alessandro Treves, Simon Stringer, Ray Kesner, Robert Robertson, Pierre Georges-François, and Shane O’Mara, and their contributions are sincerely acknowledged.

References

194


