13

Functions of Neuronal Networks in the Hippocampus and Neocortex in Memory

Edmund T. Rolls

I. Functions of the Primate Hippocampus in Memory

It is known that damage to certain regions of the temporal lobe in humans produces anterograde amnesia evident as a major deficit following the damage in learning to recognize new stimuli (Scoville and Milner, 1957; Milner, 1972; Squire, 1986; this volume, Chapter 12). The anterograde amnesia has been attributed to damage to the hippocampus, which is within the temporal lobe, and to its associated pathways such as the fornix (Scoville and Milner, 1957; Milner, 1972; Gaffan, 1974, 1977), but this has been questioned, and instead it has been suggested that damage to both the hippocampus and the amygdala is crucial in producing anterograde amnesia, in that combined but not separate damage to the hippocampus and amygdala produced severe difficulty with visual and tactual recognition tasks in the monkey (Mishkin, 1978, 1989, 1982; Murray and Mishkin, 1984, 1985). In investigations of the particular aspects of memory for which the hippocampus may be essential, it has been shown that monkeys with damage to the hippocampus–fornical system have a learning deficit on memory tasks that require them to make associations between a stimulus—for example, a picture—and a spatial motor response such as touching one part of a screen (Gaffan, 1985; Rupniak and Gaffan, 1987), and are also impaired on memory tasks that require combinations of stimulus attributes with their locations in space to be processed together, such as memory not only for which object was shown but where it was shown (Gaffan and Saunders, 1985). Further, humans with right temporal-lobe damage are also impaired in conditional spatial response and object-place memory tasks (Petrides, 1985; Smith and Milner, 1981).

campus in memory, the activity of 1510 single hippocampal neurons was recorded in thesis monkeys learning and performing these memory tasks known to be impaired by damage to the hippocampus or fornix (Rolls et al., 1989; Miyashita et al., 1989; Cahusac et al., 1989).

In an object-place memory task in which the monkey had to remember not only which object had been seen in the previous 7–15 trials but also the position in which it had appeared on a video monitor, neurons were found that responded differentially depending on which place on the monitor screen objects were shown (Rolls et al., 1989). These neurons comprised 9.4% of the population recorded. It is notable that these neurons responded to particular positions in space (whereas “place” cells in the rat respond when the rat is in a particular place; O’Keefe, 1983). In addition, 2.4% of neurons responded more to a stimulus the first time it was shown in a particular position than the second time. These neurons thus responded to a combination of information about the stimulus being shown and about position in space, for only by responding to a combination of this information could the neurons respond only when a stimulus was shown for the first time in a certain position in space.

In tasks in which the monkeys had to acquire associations between visual stimuli and spatial responses, 14.2% of the neurons responded to particular combinations of stimuli and responses (Miyashita et al., 1989). For example, in a task in which the monkey had to perform one response (touching a screen three times) when one visual stimulus was shown, but had to perform a withholding response for 3 sec to obtain reward when a different stimulus was shown (Gaffan, 1985), 9.2% of the neurons responded to one of the stimuli if it was linked to one of the responses in this task. The same neurons typically did not respond if the same stimuli or the same responses were used in different tasks, or if other stimuli were associated with the same responses in this task. Thus these neurons responded to a combination of a particular stimulus with a particular spatial motor response (Miyashita et al., 1988).

It was possible to study the activity of 41 hippocampal neurons while the monkeys learned new associations between visual stimuli and spatial responses. In some cases it was possible to show that the activity of these neurons became modified during this learning (Cahusac et al., 1986, 1989). Interestingly, 33% of the neurons that altered their responses during this learning showed a sustained differential response, but 67% of the neurons differentiated between the stimuli only at or just before the monkey learned the task, and stopped differentiating after 5–10 more trials. This is consistent with the possibility discussed below that the neurons that show large sustained differential responses inhibit the other neurons that show transient modification, so that as a result of competition not all neurons are allocated to one stimulus–spatial response association.
responses related to certain types of memory. One type of memory involves complex conjunctions of environmental information—for example, when information about position in space (perhaps reflecting information from the parietal cortex) must be memorized in conjunction with what that object is (perhaps reflecting information from the temporal lobe visual areas), so that where a particular object was seen in space can be remembered. The hippocampus is ideally placed anatomically for detecting such conjunctions, in that it receives highly processed information from association areas such as the parietal cortex (conveying information about position in space), the inferior temporal visual cortex (conveying a visual specification of an object), and the superior temporal cortex (conveying an auditory specification of a stimulus) (Van Hoesen, 1982) (see Fig. 5). The positions of stimuli in space may be represented by the firing of hippocampal neurons as described above so that conjunctions of, for example, objects and their position can be formed. It is also suggested that one neurophysiological mechanism by which “place” cells in the rat (see O’Keefe, 1983) may be formed is by conjunction learning of sets of simultaneously occurring stimuli in different parts of space, each set of which defines a place.

II. Computational Theory of the Hippocampus

A possible theoretical basis for these results, and in particular how the hippocampus may perform the conjunction or combination learning just described, is now considered.

A schematic diagram of the connections of the hippocampus is shown in Fig. 1. One feature is that there is a sequence of stages, in each of which there is a major set of input axons that connect via a form of matrix with the output neurons of that stage. The type of computation that could be performed by one of these stages is considered first.

The perforant path connections with the dentate granule cells may be taken as an example. A version of this represented as a simplified matrix is shown in Fig. 2. Although the perforant path makes one set of synapses with the output neurons in a form of matrix, the matrix is clearly very different from an association matrix memory, in that in the hippocampal system there is no unconditioned stimulus that forces the output neurons to fire (see Rolls, 1987). Nor is there for each output cell a climbing fiber that acts as a teacher as in the cerebellum (see Ito, 1984; see Chapters 9 and 10 in this volume). In the hippocampal circuit there is apparently no teacher—that is, this appears to be an example of an unsupervised learning system. The following describes one mode of operation for such a network. Later, properties of the hippocampus that

Figure 1.
(a) Representation of connections within the hippocampus. Inputs reach the hippocampus through the perforant path (1), which makes synapses with the dendrites of the dentate granule cells and also with the apical dendrites of the CA3 pyramidal cells. The dentate granule cells project via the mossy fibers (2) to the CA3 pyramidal cells. The well-developed recurrent collateral system of the CA3 cells is indicated. The CA3 pyramidal cells project via the Schaffer collaterals (3) to the CA1 pyramidal cells, which in turn have connections (4) to the subiculum. (b) Schematic representation of the connections of the hippocampus, showing also that the cerebral cortex (neocortex) is connected to the hippocampus via the parahippocampal gyrus and entorhinal cortex, and that the hippocampus projects back to the neocortex via the subiculum, entorhinal cortex, and parahippocampal gyrus.
A simulation of the operation of such a matrix is shown in Fig. 3. It is shown that the network effectively selects different output neurons to respond to different combinations of active input horizontal lines. It thus performs a type of classification in which different complex input patterns are encoded economically onto a few output lines. It should be noted that this classification finds natural clusters in the input events; orthogonalizes the classes in that overlap in input events can become coded onto output neurons with less overlap, and in that many active input lines may be coded onto few active output lines; and does not allocate neurons to events that never occur (cf. Marr, 1970, 1971; Rumelhart and Zipser, 1986; Grossberg, 1982). It may be noted that there is no special correspondence between the input pattern and which output lines are selected. It is thus not useful for any associative mapping between an input and an output event, and is thus different from linear associative matrix memories (Rolls, 1987). Instead, this type of matrix finds associations or correlations between input events (which are expressed as sets of simultaneously active horizontal input lines or axons), allocates output neurons to reflect the complex event, and stores the required association between the input lines onto the few output neurons activated by each complex input event.

There is some evidence that in the hippocampus the synapses between inactive axons and active output neurons become weaker (see McNaughton, 1983; Levy, 1985). The effect of this in the learning system described would be to facilitate accurate and rapid classification, in that weakening synapses onto a postsynaptic neuron from axons that are not active when it is strongly activated would reduce the probability that it will respond to a stimulus that must be placed into a different class. It is also of interest that (postulate 4) it is not physiologically unreasonable that the total synaptic strength onto a postsynaptic neuron is somewhat fixed (Levy and Desmond, 1985).

Another feature of hippocampal circuitry is the mossy-fiber system, which connects the granule cells of the dentate gyrus to the CA3 pyramidal cells of the hippocampus. Each mossy fiber forms approximately 10 "mosses," in which there are dendrites of perhaps five different CA3 pyramidal cells. Thus each dentate granule cell may contact approximately 50 CA3 pyramidal cells (in the mouse; see Braitenberg and Schuz, 1983). In the rat, each mossy fiber forms approximately 14 "mosses" or contacts with CA3 cells, there are 1 \times 10^6 dentate granule cells and thus 14 \times 10^6 mosses onto 0.18 \times 10^6 CA3 cells (D. Amalric, personal communication), and thus each CA3 pyramidal cell may be contacted by approximately 78 dentate granule cells. This means that (in the rat) the probability that a CA3 cell is contacted by a given dentate granule cell is 78 synapses/10^6 granule cells = 0.000078. These mossy-fiber synapses are very large, presumably because with such a relatively small number

Consider a matrix memory of the form shown in Fig. 2 in which the strengths of the synapses between horizontal axons and the vertical dendrites are initially random (postulate 1). Because of these random initial synaptic weights, different input patterns on the horizontal axons will tend to activate different output neurons (in this case, granule cells). The tendency for each pattern to select or activate different neurons can then be enhanced by providing mutual inhibition between the output neurons, to prevent too many neurons responding to that stimulus (postulate 2). This competitive interaction can be viewed as enhancing the differences in the firing rates of the output cells [cf. the contrast enhancement described by Grossberg (1982)]. Synaptic modification then occurs according to the rules of long-term potentiation in the hippocampus, namely, that synapses between active afferent axons and strongly activated postsynaptic neurons increase in strength (see McNaughton, 1983; Levy, 1985; Kelso et al., 1986; Wigstrom et al., 1986; see Chapter 14 of this volume) (postulate 3). The effect of this modification is that the next time the same stimulus is repeated, the neuron responds more (because of the strengthened synapses), more inhibition of other neurons occurs, and then further modification to produce even greater selectivity is produced. The response of the system thus climbs over repeated iterations. One effect of this observed in simulations is that a few neurons then obtain such strong synaptic weights that almost any stimulus that has any input to that neuron will succeed in activating it. The solution to this is to limit the total synaptic weight that each output (postsynaptic) neuron can receive. In simulations this is performed by normalizing the sum of the synaptic weights on each neuron to a constant (e.g., 1.0) (cf. von der Malsburg, 1973; Rumelhart and Zipser, 1986). This has the effect

![Figure 2.](https://example.com/figure2.png)

*Figure 2.*
A matrix for competitive learning in which the input stimuli are presented along the rows of the input axons (a), which make modifiable synapses (s), with the dendrites of the output neurons, which form the columns (d) of the matrix.
One effect that can be achieved by this low probability of contact of a particular dentate granule cell with a particular pyramidal cell is pattern separation. This is achieved in the following way. Consider a pattern of firing present over a set of dentate granule cells. The probability that any two CA3 pyramidal cells receive synapses from a similar subset of the dentate granule cells is very low (because of the low probability of contact of any one dentate granule cell with a pyramidal cell), so that each CA3 pyramidal cell is influenced by a very different subset of the active dentate granule cells. Thus each pyramidal cell effectively samples a very small subset of the active granule cells, and it is therefore likely that each CA3 pyramidal cell will respond differently to the others, so that in this way pattern separation is achieved. (The effect is similar to codon formation described in other contexts by Marr (1970).) With modifiability of the mossy-fiber synapses, CA3 neurons learn to respond to just those subsets of activity that do occur in dentate granule cells. Moreover, because of the low probability of contact, and because of the competition between the CA3 neurons, the patterns that occur are evenly distributed over different ensembles of CA3 neurons. This pattern separation effect can be seen in Fig. 4. (It may be noted that even if competition does not operate in this system to increase orthogonality, then the low probability of connections just described would nevertheless mean that the hippocampus could operate to produce relatively orthogonal representations.)

It is notable that in addition to the mossy fiber inputs, the CA3 pyramidal cells also receive inputs directly from perforant path fibers...
Hippocampal stimulation. Conventions as in Fig. 3. The dentate gyrus is shown as a competition matrix at the left, receiving input stimuli from the perforant path. The vertical dendrites of the CA3 pyramidal cells extend throughout the three submatrices shown in the middle. The middle submatrix receives the output of the dentate granule cells via the mossy fibers with potentially powerful synapses and a low contact probability, and operates as a competition matrix. Pattern separation can be seen to operate in that input vectors are converted into output vectors with many

(see Figs. 1 and 2). This is not a sparse projection, in that each pyramidal cell may receive on the order of 2300 such synapses. (This is calculated using the evidence that of 15 mm of total dendritic length with 10,000 spines, approximately 3.5 mm (range 2.5-4.5 mm) is in the lacunosum molecular and thus receives inputs from the perforant path (D. G. Amaral, personal communication).] What would be the effect of this input together with the very sparse, but strong, synapses from the mossy fibers? One effect is that the mossy-fiber input would cause the pattern of synapses (considered as a vector) on each pyramidal cell to point in a different direction in a multidimensional space. However, the precise direction in that multidimensional space could not be well specified by the relatively small number of mossy-fiber synapses onto each CA3 pyramidal cell. However, once pointed to that part of space by the mossy fibers, a particular cell would show cooperative Hebbian learning between its activation by the mossy input and the direct perforant path input, allowing the direct input to come by learning to specify the exact direction of that cell in multidimensional space much more effectively than by the coarse mossy-fiber input alone. This effect can be seen in Fig. 4. The relative weighting in this simulation was that the mossy-fiber input had an effect on each neuron that was five times greater than that of the direct perforant path input. Thus it is suggested that the combination of the sparse mossy-fiber input and the direct perforant path input is to achieve pattern separation, and at the same time to allow the response of the neuron to be determined not just by the sparse mossy-fiber input, but much more finely by making use in addition of the direct perforant path input.

An additional feature of the hippocampus, which is developed in the CA3 pyramidal cells in particular, is the presence of strong recurrent collaterals, which return from the output of the matrix to cross over the neurons of the matrix, as shown in Figs. 1 and 2. This anatomy immediately suggests that this is an autoassociation matrix. The effect of such recurrent collaterals is to make that part of the matrix into an autoassociation (or autocorrelation) matrix. The autoassociation arises because the output of the matrix, expressed as the firing rate of the CA3

and in that different output vectors are produced for even quite similar input vectors. The upper CA3 submatrix operates as a competition matrix with a direct perforant path input. The lower CA3 submatrix operates as an autoassociation matrix (formed by the recurrent collaterals). The output of the CA3 cells (summed vertically up and down the dendrite) is then used as the input (via the Schaffer collaterals) to the CA1 cells, which operate as a competition matrix. The states of the matrices after 200 presentations of the same set of stimuli used for Fig. 3 are shown. One point demonstrated is that two very similar stimuli, overlap six A in Fig. 4a, and overlap six B in Fig. 4b, produce output vectors of CA1 that
pyramidal cells, is fed back along the horizontally running axons so that the pattern of activity in this part of the matrix (the CA3 pyramidal cells) is associated with itself (see, e.g., Kohonen et al., 1981; Rolls, 1987a). It can be noted here that for this suggestion to be the case, the synapses of the recurrent collaterals would have to be modifiable, and the modification rule would require alteration of synaptic strength when both the presynaptic fiber and the postsynaptic dendrite were strongly activated. Further, the probability of contact of the neurons in the autoassociation matrix must not be very low if it is to operate usefully (see Marr, 1971). Given that the region of the CA3 cell dendrites on which the recurrent collaterals synapse is long (approximately 11.5 mm), and that the total dendritic length is approximately 15 mm and has approximately 10,000 spines (D. G. Amaral, personal communication), approximately 7700 synapses per CA3 pyramidal cell could be devoted to recurrent collaterals, which with 180,000 CA3 neurons in the rat makes the probability of contact between the CA3 neurons 0.043. This is high enough for the system to operate usefully as an autoassociation memory (see Marr, 1971). It is remarkable that the contact probability is so high, and also that the CA3 recurrent collateral axons travel so widely in all directions that they can potentially come close to almost all other CA3 neurons (D. G. Amaral, personal communication).

The importance of the autoassociation performed by this part of the matrix is that it forms a recognition memory, with all the advantageous emergent properties of a matrix memory, such as completion, generalization, and graceful degradation (see Kohonen et al., 1977, 1981; Kohonen, 1984; Rolls, 1987). One property that is particularly relevant here is completion, in that if part of a stimulus (or event) occurs, then the autoassociation part of the matrix completes that event. Completion may operate particularly effectively here, because it operates after the granule-cell stage, which will reduce the proportion of neurons firing to represent an input event to a low number partly because of the low probability of contact of the granule cells with the CA3 pyramidal cells. It is under these conditions that the simple autocorrelation effect can reconstruct the whole of one pattern without interference, which would arise if too high a proportion of the input neurons was active. It is of interest that a scheme of this type, although expressed in a different way to the autoassociation matrix formulation, was proposed by Marr in 1971. Another effect of the autoassociation matrix is that patterns of activity that are not similar to those already learned by this type of recognition memory are lost, so that noisy patterns can be cleaned up by the autoassociation matrix. It is further notable that these completion and cleaning-up processes benefit from several iterations (repeated cycles) of the autoassociation feedback effect. It has been suggested by B. McNaughton (personal communication, 1987) that one function of hippocampal theta activity may be to allow this autoassociation effect to

duced by the recurrent collaterals to cycle for several iterations (in a period of approximately 50 msec), and then to stop, so that the system can operate again with maximal sensitivity to new inputs received on the mossy fiber and perforant path systems by the CA3 cells.

The CA1 pyramidal cells that receive from the CA3 cells are considered to form a further stage of competitive learning, which has the effect of further classification of signals received, perhaps enabled by the pattern of connections within the hippocampus to form these classifications over inputs received from any part of the association neocortex. The firing of the CA1 cells would thus achieve a much more economical and orthogonal classification of signals than that present in the perforant path input to the hippocampus. These signals are then returned to the association neocortex via the subiculum, entorhinal cortex, and parahippocampal gyrus, as indicated in Figs. 1 and 2. It is suggested below that one role that these economical (in terms of the number of activated fibers) and relatively orthogonal signals play in neocortical function is to guide information storage or consolidation in the neocortex.

It may be noted that multilayer networks (such as the hippocampus) can potentially solve classes of problems that cannot be solved in principle by single-layer nets (Rumelhart and Zipser, 1986). This is because subcategories formed in an early stage of processing can enable a later stage to find solutions or categories that are not linearly separable in the input information space.

Having considered the computational theory of how the hippocampal circuitry may function, we can now turn to a systems-level analysis, in which the inputs and outputs of the hippocampus are considered, and the function performed by the hippocampus in relation to overall brain function can be formulated.

III. Systems-Level Theory of Hippocampal Function

First the anatomical connections of the primate hippocampus with the rest of the brain will be considered, in order to provide a basis for considering how the computational ability of the hippocampus could be used by the rest of the brain.

The hippocampus receives inputs by two main routes, the entorhinal cortex and the fimbria/fornix. The entorhinal cortex (area 28) provides it with extensive inputs from the neocortex (see Fig. 5). Thus all temporal neocortical areas project to area 35, the perirhinal cortex, or to area TF-TH, in the parahippocampal gyrus, which in turn projects to the entorhinal cortex (Van Hoesen and Pandya, 1975a, b; Van Hoesen, 1982; Amaral, 1987). The parietal cortex (area 7) projects to area TF-TH...
cortex, areas 12 and 13, projects directly to the entorhinal cortex (Van Hoesen et al., 1975). In addition, the entorhinal cortex receives inputs from the amygdala. The entorhinal cortex itself projects via the perforant path to reach primarily the dentate granule cells of the hippocampus proper. Thus by these routes the hippocampus receives information after it has been highly processed through the temporal, parietal, and frontal cortices. It must thus be of great importance for hippocampal function in the primate that by its main input, the perforant path, it receives information from the highest parts of the neocortex. There are also inputs to the hippocampus via the fimbria/fornix from the cholinergic cells of the medial septum and the adjoining limb of the diagonal band of Broca. The hippocampus also receives a noradrenergic input from the locus coeruleus, and a 5-hydroxytryptamine (5-HT) input from the median raphe nucleus.

A major output of the hippocampus arises from the hippocampal pyramidal cells, and projects back via the subiculum to the entorhinal cortex, which in turn has connections back to area TF–TH, which in turn projects back to the neocortical areas from which it receives inputs (Van Hoesen, 1982) (see Figs. 5 and 1). Thus the hippocampus can potentially influence the neocortical regions from which it receives inputs. This is the pathway suggested as being involved in guiding memory storage in the neocortex. A second efferent projection of the hippocampal system reaches the subiculum from the CA1 pyramidal cells, and travels via the fimbria and (postcommissural) fornix to the anterior thalamus, and to the mammillary bodies, which in turn project to the anterior thalamus. The anterior thalamus in turn projects into the cingulate cortex, which itself has connections to the supplementary motor cortex, providing a potential route for the hippocampus to influence motor output (Van Hoesen, 1982). It is suggested that functions of the hippocampus in, for example, conditional spatial response learning utilize this output path to the motor system.

The connections of the hippocampus with other parts of the brain, and the internal connections and synaptic modifiability described above, suggest that the hippocampus should be able to detect, and classify onto a few specifically responding neurons, specific conjunctions of complex (cortically processed) events, such as a particular object (presumably reflecting temporal lobe visual processing) that has appeared in a particular position in space (presumably reflecting parietal input). Another example might be that a particular stimulus should be associated with a particular spatial motor response. It has been shown above that this is the type of quite specific information that comes to activate different hippocampal neurons while monkeys are performing object–place memory and conditional spatial response learning tasks. Indeed, the neurophysiological findings described above provide evidence that supports the model of hippocampal function just described. The model is also supported by the evidence that during learning of conditional spatial responses some hippocampal neurons start, but then stop, showing differential responses to the different stimuli, consistent with competitive interactions between hippocampal neurons during learning, so that only some hippocampal neurons become allocated to any one learned event or contingency (see above).

The analyses above have shown that the hippocampus receives from high-order areas of association cortex; is able by virtue of the large number of synapses on its dendrites to detect conjunctions of events even when these are widely separated in information space, with their origin from quite different cortical areas; allocates neurons to code efficiently for these conjunctions probably using a competitive learning mechanism; and has connections back to the neocortical areas from which it receives, as well as to subcortical structures via the fimbria/fornix system. What could be achieved by this system? It appears that the long-term plan is for the hippocampus to be involved in the learning of conditional spatial responses and object–place memories. The hippocampus is therefore ideally suited for tasks that require the learning of spatial and object–place memories. It is also well suited for the active exploration of the environment, which is essential for learning these tasks. The hippocampus is therefore well suited for the learning of conditional spatial responses and object–place memories. The hippocampus is therefore ideally suited for tasks that require the learning of spatial and object–place memories. It is also well suited for the active exploration of the environment, which is essential for learning these tasks.
mans, in that damage to the hippocampus in humans does not necessarily result in major retrograde amnesia (Squire, 1986; this volume, chapter 12). On the other hand, the hippocampus does appear to be necessary for the storage of certain types of information (characterized by the description declarative, or knowing that, as contrasted with procedural, or knowing how). How could the hippocampus then be involved in the storage of information?

The suggestion that is made on the basis of these and the other findings described above is that the hippocampus is specialized to detect the best way in which to store information, and then by the return paths to the neocortex directs memory storage there. Clearly the hippocampus, with its large number of synapses on each neuron and its potentiation type of learning, is able to detect when there is coherence (i.e., conjunctive activation of arbitrary sets of its input fibers), and is able, as indicated both theoretically and by recordings made in the behaving monkey, to allocate neurons to economically (i.e., with relatively few neurons active) code for each complex input event. Such neurons could then represent an efficient way in which to store information, in that redundancy would effectively have been removed from the input signal. In a sense, the optimal way in which to build high-level feature analyzers could be determined by the hippocampus. It should be noted that this theory is not inconsistent with the possibility that the hippocampus provides a working memory, in that in the present theory the hippocampus sets up a representation using Hebbian learning, which is useful in determining how information can best be stored in the neocortex. [The representation found by the hippocampus could provide a useful working memory (see Olton, 1983), and indeed in the object-place memory task described above the object and place combinations formed onto single hippocampal neurons would provide a useful working memory. It may be that by understanding the operations performed by the hippocampus at the neuronal network level, it can be seen how the hippocampus could contribute to several functions that are not necessarily inconsistent.]

The question then arises of where the long-term storage occurs, and how it may be directed by the hippocampus. Now the hippocampus is reciprocally connected via the subiculum and entorhinal cortex with the parahippocampal gyrus, which in turn is reciprocally connected with many high-order areas of association neocortex (see Fig. 5). It is therefore possible that the actual storage takes place in the parahippocampal gyrus, and that this might be particularly appropriate for multimodal memories. However, having detected that, for example, a visual stimulus is regularly associated with an event in another modality such as a response, it might be useful to direct the unimodal representation of that visual image, so that it is stored efficiently and can provide a useful input to the multi-modal conjunction store. Thus it is suggested that return pathways (for example, via the parahippocampal gyrus) to unimodal cortex (for example, inferior temporal cortex, area TE) might be used to direct unimodal storage, too, by contributing to detection of the most economical way in which to store representations of stimuli.

The question of how the hippocampal output is used by the neocortex will be considered next. Given that the hippocampal output returns to the neocortex, a theory of backprojections in the neocortex will be needed. This is developed next. By way of introduction to this, it may be noted that which particular hippocampal neurons happen to represent a complex input event is not determined by any teacher or forcing (unconditioned) stimulus. Thus the neocortex must be able to utilize the signal rather cleverly. One possibility is that any neocortical neuron with a number of afferents active at the same time that hippocampal return fibers in its vicinity are active modifies its responses, so that it comes to respond better to those afferents the next time they occur. This learning by the cortex would involve a Hebb-like learning mechanism. It may be noted that one function served by what are thus in effect backprojections from the hippocampus is some guidance for or supervision of neocortical learning. It is a problem of unsupervised learning systems that they can detect local conjunctures efficiently, but that these are not necessarily those of most use to the whole system. It is proposed that it is exactly this problem that the hippocampus helps to solve, by detecting useful conjunctures globally (i.e., over the whole of information space), and then directing storage locally at earlier stages of processing so that filters are built locally that provide representations of input stimuli that are useful for later processing.

IV. Theoretical Significance of Backprojections in the Neocortex

The forward and backward projections that will be considered are shown in Fig. 6 [for further anatomical information see Peters and Jones (1984)]. In primary sensory cortical areas, the main extrinsic "forward" input is from the thalamus and ends in layer 4, where synapses are formed onto spiny stellate cells. These in turn project heavily onto pyramidal cells in layers 3 and 2, which in turn send projections forward to terminate strongly in layer 4 of the next cortical layer [on small pyramidal cells in layer 4 or on the basal dendrites of the layer-2 and -3 (superficial) pyramidal cells]. (Although the forward afferents end strongly in layer 4, the forward afferents have some synapses also onto the basal dendrites of the layer-2 pyramidal cells, as well as onto layer-6 pyramidal cells and inhibitory interneurons). Inputs reach the layer-5 (deep) pyramidal cells from the pyramidal cells in layers 2 and 3 (Martin, 1984), and it is the
Cortical connections

1
2

Amygdala
Hippocampus

Figure 6.
Schematic diagram of forward and backward projections in the neocortex. The superficial pyramidal cells (triangles) in layers 2 and 3 project forward to terminate in layer 4 of the next cortical area. The deep pyramidal cells in the next area project back (mainly) to layer 1 of the preceding cortical area, in which there are apical dendrites of pyramidal cells. The hippocampus and amygdala also are the source of backprojections that end (mainly) in layer 1. Spiny stellate cells are represented by small circles in layer 4. See text for further details.

preceding cortical area (see Fig. 6), where there are apical dendrites of pyramidal cells. There are few current theories about the functions subserved by the cortico-cortical backprojections, even though there are almost as many backprojecting as forward-projecting axons. It is important to note that in addition to the axons and their terminals in layer 1 from the succeeding cortical stage, there are also axons and terminals in layer 1 in many stages of the cortical hierarchy from the amygdala and (via the subiculum, entorhinal cortex, and parahippocampal cortex) from the hippocampal formation (see Fig. 6) (Van Hoesen, 1981; Turner, 1981; Amaral and Price, 1984; Amaral, 1986, 1987). The amygdala and hippocampus are stages of information processing at which the different sensory modalities (such as vision, hearing, touch, and taste for the amygdala) are brought together, so that correlations between inputs in different modalities can be detected in these regions, but not at prior cortical processing stages in each modality, as these cortical processing stages are unimodal. As a result of bringing together the two modalities, significant correspondences between the two modalities can be detected. One example might be that a particular visual stimulus is associated with the taste of food. Another example might be that another visual stimulus is associated with painful touch. Thus at these stages of processing, but not before, the significance of, for example, visual and auditory stimuli can be detected and signalled, and sending this information back to the neocortex thus can provide a signal that indicates to the cortex that information should be stored, but even more than this, provides an orthogonal signal that could help the neocortex to store the information efficiently.

The way in which backprojections could assist learning in the cortex can be considered using the architecture shown in Fig. 7. The (forward) input stimulus occurs as a vector applied to (layer-3) cortical pyramidal cells through modifiable synapses in the standard way for a competitive net. (If it is a primary cortical area, the input stimulus is at least partly relayed through spiny stellate cells, which may help to normalize and orthogonalize the input patterns in a preliminary way before the patterns are applied to the layer-3 pyramidal cells. If it is a nonprimary cortical area, the cortico-cortical forward axons may end more strongly on the basal dendrites of neurons in the superficial cortical layers.) The lower set of synapses on the pyramidal cells would then start by competitive learning to set up representations on the lower parts of these neurons, which would represent correlations in the input information space and could be said to correspond to features in the input information space, where a feature is defined simply as the representation of a correlation in the input information space.

Consider now the application of one of the (forward) input stimulus vectors with the conjunctive application of a pattern vector via the backprojection axons with terminals in layer 1. Given that all the synapses in the matrix start with random weights, some of the pyramidal cells will by chance be strongly influenced both by the (forward) input stim-

Figure 7.
The architecture used to simulate the properties of backprojections. The forward-projected ("input stimulus") and backprojected axons make Hebb modifiable synapses onto the same set of vertical dendrites, which
ulus and by the backprojecting vector. These strongly activated neurons will then compete with each other as in a standard competitive net, to produce contrast enhancement of their firing patterns. [The relatively short-range (50 μm) excitatory operations produced by bipolar and double bouquet cells, together with more widespread (300–500 μm) recurrent lateral inhibition produced by the smooth nonpyramidal cells and perhaps the basket cells, may be part of the mechanism of this competitive interaction.] Next, Hebbian learning takes place as in a competitive net, with the addition that not only are the synapses between forward-projecting axons and active postsynaptic neurons modified, but also the synapses in layer 1 between the backward projecting axons and the (same) active postsynaptic neurons are modified.

This functional architecture has the following properties. First, the backprojections, which are assumed to be relatively information-rich and orthogonal to each other as a result of the conjunctions formed by the hippocampus, amygdala, or next cortical stage, help the neurons to learn to respond differently to (and thus to separate) input stimuli (on the forward projection lines) even when the stimuli are very similar. This is illustrated in the simulation shown in Fig. 8, in which it is shown that input stimuli that overlap even in six positions out of eight can be easily learned as separate if presented conjunctively with different orthogonal backprojecting "tutors." [For a similar idea on the guidance of a competitive learning system see Rumelhart and Zipser (1986).]

In the neocortex, the backprojecting tutors can be of two types. One originates from the amygdala and hippocampus, and, by benefiting from cross-modal comparison, can provide an orthogonal backprojected vector. Moreover, this backprojection may only be activated if the multimodal areas detect that the visual stimulus is significant, because (for example) it is associated with a pleasant taste. This provides one way in which guidance can be provided for a competitive learning system as to what it should learn, so that it does not attempt to lay down representations of all incoming sensory information. Another way for this important function to be achieved is by activation of neurons that "strobe" the cortex when new or significant stimuli are shown. The cholinergic system originating in the basal forebrain (which itself receives information from the amygdala), and the noradrenergic input to layer 1 of the cortex from the locus coeruleus may also contribute to this function (see Rolls, 1987; Bear and Singer, 1986). However, in that there are relatively few neurons in these systems, it is suggested that these projections only provide a simple "strobe," rather than carrying pattern-specific information to guide how information is consolidated (Rolls, 1987). The second type of backprojection is that from the next cortical area in the hierarchy. This operates in the same manner, and because it is a competitive system, is able to further categorize or orthogonalize more orthogonal representations as tutors to the preceding stage, to effectively build at the preceding stage better filters for the diagnosis of the categories being found at the next stage.

A second property of this architecture is that if only the tutor is presented, then the neurons originally activated by the forward projecting input stimuli are activated. This occurs because the synapses from the backprojecting axons onto the pyramidal cells have been modified only where there was conjunctive forward and backprojected activity during learning. This thus provides a mechanism for recall. A simulation of this is shown in Fig. 8. Consider the situation when in the visual system the sight of food is forward-projected onto pyramidal cells, and conjunctively there is a backprojected representation of the taste of the food. Neurons that have conjunctive inputs from these two stimuli set up representations of both, so that later if only the taste representation is backprojected, then the visual neurons originally activated by the sight of that food will be activated. In this way many of the low-level details of the original visual stimulus can be recalled. Evidence that during recall relatively early cortical processing stages are activated comes from cortical blood flow studies in humans, in which it has been found, for example, that quite early visual cortical association areas are activated during recall of visual (but not auditory) information (Roland and Friberg, 1985; Roland et al., 1980).

A third property of this architecture is that attention could operate from higher to lower levels to selectively facilitate only certain pyramidal cells by using the backprojections. Indeed, the backprojections described could produce many of the "top-down" influences that are common in perception. A fourth property is that semantic priming could operate by using the backprojecting neurons to provide a small activation of just those neurons in earlier stages that are appropriate for responding to that semantic category of input stimulus.

A fifth property of such a return mechanism, which on detecting a conjunction (perhaps across modalities) improved unimodal representations of the input stimuli, would be a form of positive feedback, which would result in gradually improving storage as a result of the reciprocal interactions, as the feedback effect produced a better representation at the preceding level to be fed forward, with this occurring repeatedly. This might provide a neurophysiological and computational basis for any gradient of retrograde amnesia that may occur for the period just before disruption of temporal-lobe function (Squire, 1986; this volume, chapter 12). A sixth property of the backprojections is that they would assist the stability of the preceding competitive networks by providing a relatively constant guiding signal as a result of associations made at a higher stage, for example to an unconditioned taste or somatosensory input.
Figure 8.
Simulation of neocortical backprojection learning matrix. Conventions as in Fig. 3. (a) During learning, both a forward input (chosen from the same set as used in Fig. 3) and a backprojected vector that was orthogonal to the other backprojected vectors were presented simultaneously. After 206 cycles with input stimuli A and backprojected tutor pairs chosen in random
stimuli (overlap six A and six B) produce different outputs (b and c). The learning has been guided by the backprojected tutors presented during learning. If only the tutor originally paired with input stimulus overlap six A is presented, then overlap C

requires a large number of backprojecting axons, as pattern-specific information (used to guide learning by providing a set of mutually orthogonal guidance signals, or to produce recall) must be provided by the backprojections. It also solves the de-addressing problem, for the hippocampus does not need to know exactly where in the cortex information should be stored. Instead, the backprojection signal spreads widely in layer 1, and the storage site is simply on those neurons that happen to receive strong (and precise) forward activity as well as backprojected activity. This scheme is consistent with neocortical anatomy, in that it requires the same pyramidal cell to receive both forward and (more diffuse) backprojected activity, which the arrangement of pyramidal cells with apical dendrites that extend all the way up into layer 1 achieves (see Peters and Jones, 1984). Indeed, in contrast to the relatively localized terminal distributions of forward cortico–cortical and thalamo–cortical afferents, the cortico–cortical backprojection that end in layer 1 have a much wider horizontal distribution, of up to several millimeters (Amaral, 1986). The suggestion is thus that this enables the backward projecting neurons to search over a larger number of pyramidal cells in the preceding cortical area for activity that is conjunctive with their own. It is also of interest that the theory utilizes a Hebbian learning scheme that provides for learning to occur when conjointly there is forward and backprojected input to a pyramidal cell resulting in sufficient postsynaptic activation to provide for modification of synapses that happen to be active. This provides the opportunity to make it clear that the theoretical ideas introduced here make clear predictions that can be empirically tested. For example, the theory of backprojections just proposed predicts that the backprojections in the cerebral cortex have modifiable synapses on pyramidal cells in the previous cortical area. If this were found not to be the case in empirical tests, then the theory would be rejected.

The ideas introduced here also have many theoretical implications. One is that if the backprojections are used for recall, as seems likely as discussed above (see also Roland and Friberg, 1985), then this would place severe constraints on their use for functions such as error backpropagation. Error backpropagation is an interesting and powerful algorithm in parallel distributed processing networks for setting the weights in hidden units (i.e., nodes in layers that are not input and output layers) to allow networks to learn useful mappings between input and output layers (Rumelhart et al., 1986). However, the backprojections in the architecture in which this algorithm is implemented have very precise functions in conveying error from the output layer back to the earlier, hidden, layers. It would be difficult to use the weights (synaptic strengths) from the backprojecting neurons to neurons in earlier layers both to convey the error correctly and to have the appropriate strengths for recall.

Cortical approaches to the function of the hippocampus and of backprojections in the neocortex have been considered. Theories of how the hippocampus functions and of the functions of backprojections in the neocortex have been proposed. The theories are at the level of neuronal networks, and are based partly on evidence on the fine architecture of the networks, on the rules of synaptic modifiability incorporated, and on the systems-level connections. It is suggested that this approach will be useful in the future in linking anatomical evidence on structure to physiological evidence on modifiability, understanding the global properties of the networks, and thus understanding the role of the networks in brain function and behavior.

Acknowledgments

The author has worked on some of the experiments and neuronal network modeling described here with P. Cahusac, D. Cohen, J. Feigenbaum, R. Kesner, Y. Miyashita, and H. Niki, and their collaboration is sincerely acknowledged. Discussions with David G. Amaral of the Salk Institute, La Jolla, California, were also much appreciated. This research was supported by the Medical Research Council.

References


