Annu. Rev. Psychol. 1997. 48:481–514

PSYCHOBIOLOGICAL MODELS OF HIPPOCAMPAL FUNCTION IN LEARNING AND MEMORY

Mark A. Gluck and Catherine E. Myers

Center for Molecular and Behavioral Neuroscience, Rutgers University, 197 University Avenue, Newark, New Jersey; e-mail: gluck@pavlov.rutgers.edu

KEY WORDS: hippocampus, computation, neural networks

ABSTRACT

We review current computational models of hippocampal function in learning and memory, concentrating on those that make strongest contact with psychological issues and behavioral data. Some models build upon Marr's early theories for modeling hippocampal field CA3's putative role in the fast, temporary storage of episodic memories. Other models focus on hippocampal involvement in incrementally learned associations, such as classical conditioning. More recent efforts have attempted to bring functional interpretations of the hippocampal region in closer contact with underlying anatomy and physiology. In reviewing these psychobiological models, three major themes emerge. First, computational models provide the conceptual glue to bind together data from multiple levels of analysis. Second, models serve as important tools to integrate data from both animal and human studies. Third, previous psychological models that capture important behavioral principles of memory provide an important top-down constraint for developing computational models of the neural bases of these behaviors.

CONTENTS

0084-6570/97/0201-0481$08.00

INTRODUCTION

Many models and theories have been proposed over the past few decades that attempt to characterize the role of the hippocampal region in learning and memory. Most of these theories are qualitative, consisting of a central concept or metaphor that attempts to capture the essence of hippocampal-region function. We focus here on more formal computational network models of hippocampal function in learning and memory. Such models have an advantage in that they can be rigorously tested with computer simulations and, occasionally, formal mathematical analysis.

Given the breadth and diversity of current hippocampal models, we concentrate on just that subset of hippocampal theories that make strongest contact with psychological issues and data from behavioral studies of learning and memory. Given this psychobiological perspective, we omit more physiological models that make less contact with behavioral aspects of learning and memory. Among the theories that do address observable memory behaviors, we emphasize those that relate most strongly to traditional theories and models within the literature.

Our purpose in the first part of the review is to provide a general understanding of the aims, successes, and limitations of the computational approach to understanding hippocampal function in learning and memory behavior. The emphasis is on describing the spirit and behavior of the models, rather than on their exact mathematical underpinnings. A few mathematical equations are given where critical to this description. For a full exposition on implementation details, see the original journal articles.

The remainder of the review is organized as follows. We present a brief summary of the major points of hippocampal anatomy and a review of the empirical data on memory deficits produced by hippocampal damage in animals and human beings. We then provide some important historical background, discussing David Marr's early theories of the hippocampus as an

autoassociative memory storage device. In the section entitled "Autoassociative Models of CA3 and Episodic Memory" we show how Marr's earlier theories have influenced current computational models of hippocampal region CA3 and its role in episodic memory. Next we turn to more incremental forms of associative learning, reviewing models of conditioning and hippocampus. The next two sections illustrate how some of these models, at different levels of analysis, are beginning to converge into integrated theories incorporating a wider range of behavioral and biological detail.

THE HIPPOCAMPAL REGION IS CRITICAL FOR LEARNING AND MEMORY

The hippocampal region (Figure 1) is comprised of a group of brain structures located deep inside the brain that form part of what (in human beings) is often called the *medial-temporal lobe.* The region includes the *hippocampus* and the nearby *dentate gyrus, subiculum,* and *entorhinal cortex.* The outermost of these structures—the entorhinal cortex—receives highly processed information from the entire spectrum of sensory modalities as well as from multimodal association areas. Information flows in a roughly unidirectional fashion from the entorhinal cortex to the dentate gyrus, to the hippocampus, to the subiculum, and back to the entorhinal cortex before returning to the same sensory areas where it originally arose. In addition to this basic pathway of information flow, there are many direct connections between the structures of the region. The hippocampus also has another input and output pathway through the *fornix,* a fiber bundle connecting it with subcortical structures that provide modulation.

Hippocampal Damage Produces Amnesia in Human Beings

Damage to the hippocampal region in human beings produces a characteristic *anterograde amnesia* syndrome, which strongly impairs the acquisition of new information (Squire 1987). Human hippocampal damage can result from a variety of causes, including aneurysms to the arteries that vascularize the hippocampus, anoxia, and epileptic seizures (Zola-Morgan & Squire 1993). The hippocampal region is also among the first structures to be damaged in the course of Alzheimer's disease and normal aging (de Leon et al 1993). Damage to other related structures, such as the basal forebrain, can also cause amnesic syndromes that share features with hippocampal amnesia, presumably because such damage indirectly interferes with normal hippocampal-region processing (Volpe & Hirst 1983).

The anterograde amnesia that follows human hippocampal-region damage is most saliently characterized by an inability to acquire new *episodic* information, the kind of information about individual events and experiences that is

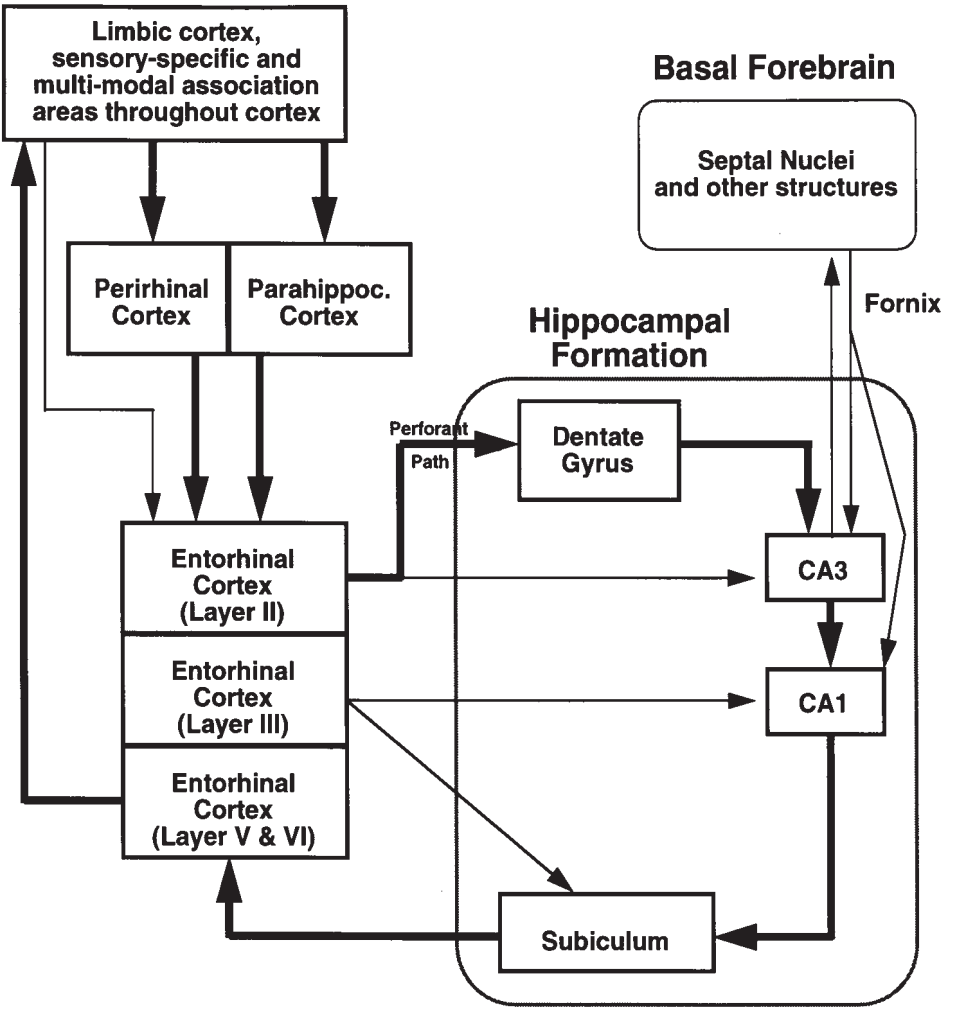

Figure 1 The structures of the hippocampal region. CA3, hippocampal field CA3; CA1, hippocampal field CA1. [Adapted from Myers et al (1996).]

accessible to conscious control. Patients with this debilitation may also show some degree of retrograde amnesia—disruption of previously acquired information—but this is usually limited to information acquired shortly before the trauma and tends to lessen in a time-graded fashion for information acquired longer ago (Squire 1987). This relationship between hippocampal damage and anterograde amnesia led to the idea that the hippocampus is a specialized memory processor needed to lay down new episodic memories.

Hippocampal Damage Produces Varied Memory Deficits in Animals

Animal models of hippocampal amnesia have had an obvious difficulty in addressing this loss of episodic information in nonverbal subjects; animals are unable to tell the experimenter directly what they can remember. However, by using indirect memory tests in which the animal is challenged to use memory of specific events, hippocampal-region damage in animals has been shown to cause learning deficits broadly similar to the episodic memory loss in human hippocampal amnesics (Eichenbaum 1992). Animal studies have also documented that certain kinds of learning capabilities do survive hippocampal-region damage. For example, the acquisition of learned responses in elementary associative conditioning tasks is largely unimpaired (Gabrieli et al 1995). Human hippocampal amnesics show similar residual learning abilities in motor-reflex conditioning, cognitive skill learning, and simple categorization tasks (Cohen 1984). All these tasks are learnable over many trials and do not require the formation of single episodic memories.

However, even the simple iterative tasks such as conditioning are disrupted in hippocampal-damaged animals if they involve additional complexities, such as requiring comparisons or configurations of multiple stimulus cues, or if attention to the experimental context is important (see section entitled "Stimulus Representation in Associative Learning"; Hirsh 1974; Rudy & Sutherland 1989, 1995).

MARR'S AUTOASSOCIATIVE MEMORY STORE

One of the earliest and most influential models of hippocampal-region processing was proposed by David Marr (1971). Starting with what was known then about hippocampal anatomy and physiology, Marr sought to infer an emergent information-processing capability. His ideas gave rise to a broad class of models, often termed *Hebb-Marr* models because they incorporate Hebb's (1949) ideas on how associations are acquired between groups of cells in the brain (McNaughton & Nadel 1990). Since Marr's original publication, new empirical data have shown that some aspects of his model are incomplete or incorrect (see e.g. Willshaw & Buckingham 1990). Nonetheless, many of the basic ideas in the Hebb-Marr model of hippocampus have withstood continuing empirical and theoretical tests and remain the basis for many current models and theories. This section reviews a generalized version of the Hebb-Marr model. Later sections describe several more current models that build on Marr's original specification.

Marr's basic idea was to distinguish separable roles in memory for the archicortex, including hippocampus, and for the neocortex. He assumed that

the chief role of neocortex was to store large complex *event memories*—broadly equivalent to what today are usually called *episodic memories*—composed of several integrated associations. For example, the event memory of a meal might include associations about the food eaten, the meal's location and time, and the company sharing it. In Marr's model, an event memory is defined as a pattern *E* of activities over a large number of neocortical cells, evoked by a particular set of sensory inputs (Figure 2*A*). Such a pattern is stored by associating its elements so that activation of some of the

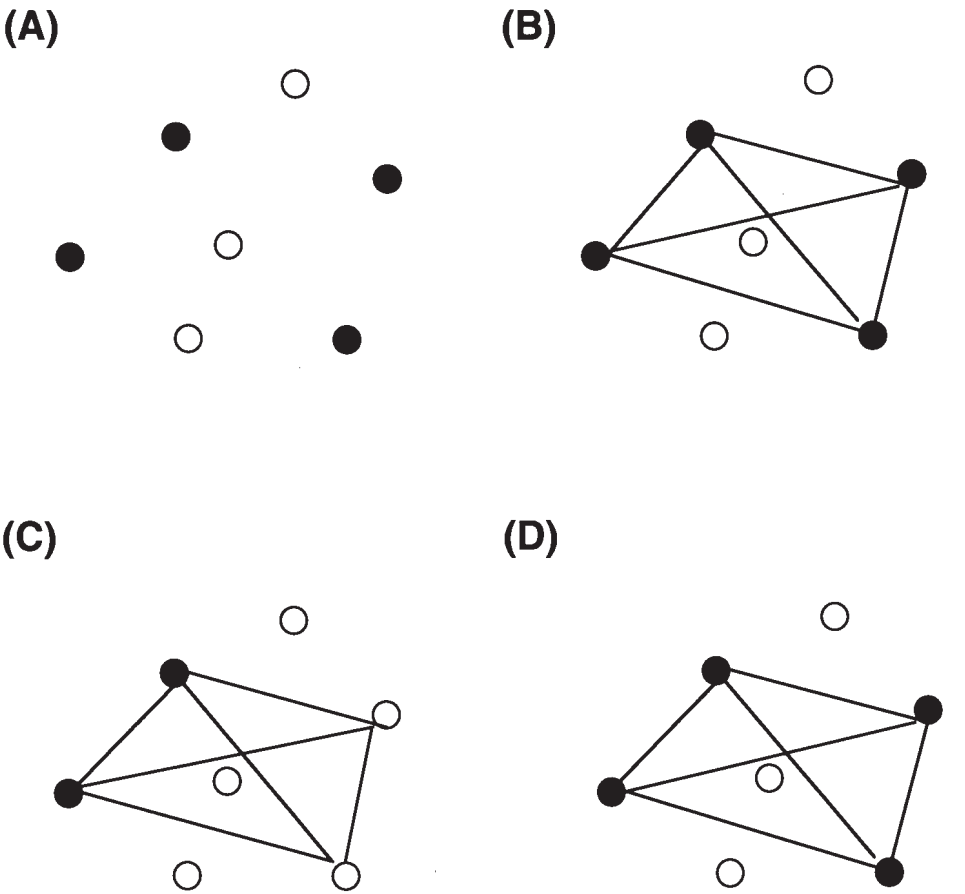

Figure 2 Storage of an event memory as a pattern of cell activations in neocortex, according to Marr's (1971) model. (*A*) Initially, the event memory simply evokes a pattern of activations (*darkened circles*) across a group of unrelated cells. (*B*) As the pattern is stored, various elements of the pattern are associated by weighted connections (*lines*). (*C*) Later, if a partial version of the original pattern is presented (*darkened circles*), activation spreads along the associations to activate the complete pattern (*D*).

cells representing elements in *E* can activate other elements in turn (Figure 2*B*). Later, if a subset of *E* is presented to the neocortex, the neocortex should be able to retrieve the full pattern *E* (Figure 2*C*). This ability is *pattern completion.* One difficulty in implementing this function in the neocortex is that a large number of very precise connections is required to associate each element in *E* with every other element in *E.* Further, the associations required to store *E* may well disrupt preexisting associations created to store other patterns with common elements. Worse, if another stored pattern *F* shares common elements with *E,* then *F* may interfere with attempted retrieval of *E:* If a subset of *E* is presented, activation will spread to these common elements, which will then begin to retrieve *F* as well as *E.* At the extreme, if many overlapping patterns are stored, an attempt to retrieve any stored pattern will result in a pattern of activation that shares elements with all stored patterns but is identical to none. This situation is called *catastrophic interference* (Hetherington 1990).

Because of this potential for interference in recall, Marr suggested that it would be useful to have a separate processor—such as the hippocampus—that could rapidly store event memories, and then allow gradual transfer of this pattern to neocortex, which would reorganize and classify this information, incorporating it with existing knowledge to reduce interference. More specifically, Marr proposed that the hippocampus is able to rapidly store new patterns, holding them in a temporary memory store, but is not able to integrate them with the larger body of existing knowledge.

Marr imagined the hippocampus as functionally consisting of two layers or groups of cells (Figure 3*A*). Inputs cause activity on the first A layer of cells, which project onto the second B layer of cells. The B cells in turn project back to the A cells. All synapses between cells are modifiable, but they are simplified to allow only binary on or off values. Similarly, cell activity is assumed to be either on or off. This network is essentially the same as that shown in Figure 2, except that cells are differentiated according to whether they directly receive external input (A cells) or not (B cells). A stored pattern can be retrieved if, when part is presented to the A cells, the evoked activity on the B cells feeds back to complete the original firing pattern on the A cells. As shown in the next section, Marr's pattern associator model forms the basis for many subsequent—and more detailed—models of hippocampal physiology and function.

AUTOASSOCIATIVE MODELS OF CA3 AND EPISODIC MEMORY

The network described by Marr is a form of *autoassociator.* An autoassociator network learns to associate an input pattern with an identical output pattern

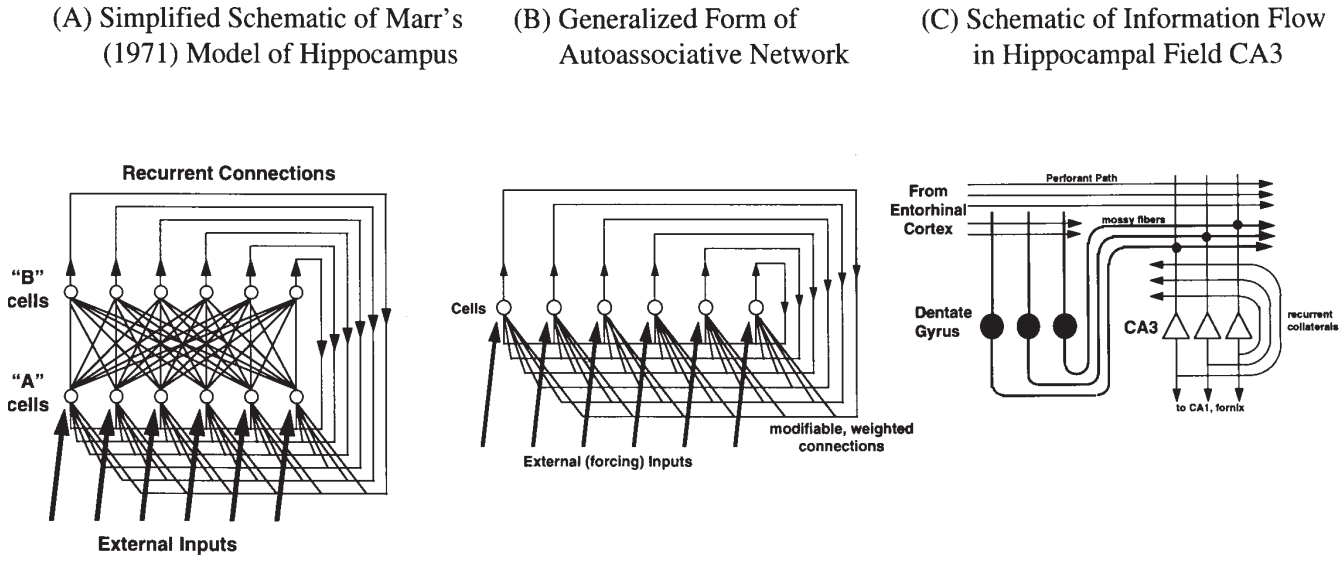

Figure 3 (*A*) Simplified schematic of Marr's (1971) model of hippocampus. Cells are either A cells, which receive direct activation from the external input (*heavy lines*), or B cells, which are driven only by A cells and afferent them in turn. Learning consists of strengthening connections between B cells and the A cells that activate them. Later, if a partial version of a stored pattern is presented to the A cells, the B cells feed back and activate the remaining A cells required to complete the stored pattern. (*B*) A generalized form of autoassociative network. There is a single layer of cells with outputs that ramify to provide feedback input to the cell layer. These synapses are weighted, and cells become active if the total weighted synaptic input exceeds a threshold (Equation 1). Patterns are stored by presenting external (forcing) input to the cells; learning then consists of weighting the synapses between all pairs of co-active cells (Equation 2). Later, if a partial version of a stored pattern is presented on the external inputs, activity spreads iteratively through the recurrent feedback connections, which activates additional cells until the entire pattern is reconstructed. (*C*) A schematic representation of information flow in hippocampal field CA3. Inputs to the pyramidal cells arrive directly from entorhinal cortex as well as indirectly via the mossy fibers from dentate gyrus. The mossy fiber afferents make sparse, presumably strong, synapses onto CA3 dendrites, and so they are putative forcing inputs to the network. CA3 pyramidal cell outputs ramify to become feedback afferents to CA3, and also exit to hippocampal field CA1 and through the fornix to other, extrahippocampal targets.

(Anderson 1977, Hinton 1989, Kohonen 1984). A general form of autoassociator is shown in Figure 2*B* and consists of a single layer of nodes receiving excitatory connections from external sources as well as from each other. Nodes are assumed to have binary states: either active or firing (represented by an output value of 1), or quiescent (represented by an output value of 0). Node *j* becomes active if the sum of inputs exceeds some firing threshold (cf Grossberg 1976, Kohonen 1984, McCulloch & Pitts 1943, Rosenblatt 1962):

$$y_j = 1 \quad \text{iff} \quad \sum_i w_{ij} y_j > \theta_j, \qquad 1.$$

else = 0.

In Equation 1, y_j is the output or activation of cell *j*, w_{ij} is the weight of the synapse on *j* from another cell *i*, and θ_j is cell *j*'s threshold. This threshold, θ_j, is then set so that *j* will become active if the weighted sum of its inputs exceeds some proportion of the total inputs active in the original pattern. Additional inhibitory processes, not shown in Figure 2*B*, may be required to determine the threshold. More complex networks may also allow continuous (real-valued) inputs and outputs, but the central ideas are the same.

A binary pattern *E* is stored in this network by presenting *E* as external input. The *n*th element of *E* is presented to the *n*th node in the network and forces that node to output the same value as that element. For this reason, the external inputs are often termed *forcing inputs*, and the one synapse each node receives from the forcing input is often called a *forcing synapse*. The network then undergoes synaptic plasticity at the feedback connections, so that synapses from active presynaptic cells have excitatory effects on other active postsynaptic cells. This can be accomplished by a Hebbian learning rule of the form:

$$w_{ij} = \alpha(y_i y_j). \qquad 2.$$

where y_i and y_j are the activities of presynaptic cell *i* and postsynaptic cell *j*, α is a constant term, and w_{ij} is the weight of the synapse between *i* and *j*. Note that synaptic mechanisms of long-term potentiation and depression (LTP and LTD) are Hebbian in nature (Levy et al 1983, McNaughton & Morris 1987). Later, if some subset of *E* is presented to the network, activity in the recurrent collaterals will iterate through the network and activate the cells needed to complete the missing parts of *E*. Thus, this network performs pattern completion.

Three Common Features of Autoassociators and Field CA3

Marr's important contribution was to conceptualize the hippocampus as an autoassociator network that performs pattern storage and retrieval. Many subsequent models have elaborated on this idea (Hasselmo 1995, Hasselmo et al

1996, McNaughton & Morris 1987, McNaughton & Nadel 1990, Rolls 1989, Treves & Rolls 1992). An autoassociator such as the one shown in Figure 3*B* has three basic requirements: (*a*) a high degree of internal recurrency among the principal cells; (*b*) strong, sparse synapses from external afferents, which could function as forcing synapses; and (*c*) plasticity at the synapses between co-active cells.

These requirements suffice to allow the functions of pattern storage, completion, and retrieval. Hippocampal field CA3 satisfies all three requirements (Figure 3*C*). First, the principal neurons of CA3—pyramidal cells—are perhaps unique in the brain for their high degree of internal recurrency: Each CA3 pyramidal may receive contact from about 4% of other pyramidals in the field, a high enough contact probability to allow autoassociation (Rolls 1989). Second, in addition to recurrent collaterals and sparse entorhinal afferents, CA3 pyramidals receive a small number of inputs from mossy fibers, containing entorhinal information that reaches CA3 via the dentate gyrus. While each CA3 pyramidal in rat may receive 12,000 synapses from recurrent collaterals and 4000 synapses from direct entorhinal afferents, it may only receive about 50 mossy fiber synapses (Treves & Rolls 1992). However, the mossy fiber synapses are very large and presumably also very strong, so that coincident activity on a relatively small number of mossy fiber synapses could activate a CA3 pyramidal (Rolls 1989). The mossy fiber synapses are thus good candidates for forcing synapses in an autoassociator (Marr 1971, McNaughton 1991, McNaughton & Morris 1987). Third, plasticity in the form of LTP has been demonstrated at the synapses of recurrent collaterals in CA3 (Bliss & Lomo 1973, Kelso et al 1986). LTP involves strengthening of synapses between coactive pre- and postsynaptic cells; this could implement Hebbian learning as defined in Equation 2 above.

In summary, CA3 seems to be a likely candidate to implement autoassociative memory in the brain. Patterns would be stored by presentation over the mossy fibers, which would force CA3 pyramidal output. Recurrent collateral synapses between coactive pyramidals would then undergo LTP to store the pattern. Later, if a partial version of that pattern is presented along the weaker entorhinal afferents, some CA3 pyramidals would become active. After several iterations of activity through the recurrent collaterals, more CA3 pyramidals would be activated until the entire stored pattern is retrieved. Additional inhibitory units are also generally assumed to allow implementation of the firing thresholds.

Autoassociative Networks Implement Hippocampal-Dependent Memory Behaviors

This type of autoassociative network can be used to implement various forms of memory, many of which are much like those that appear to be impaired

following hippocampal damage in animals and human beings. For example, autoassociative memories can create unified memories from several component features and then retrieve the entire memory from a partial input.

SEQUENCE LEARNING Many models of hippocampal function have drawn on the details of its anatomy and physiology to argue that it has the capacity for learning sequences of input patterns. These models are often generally based on the recurrent architectures of autoassociative networks: Given a partial input consisting of the present state, an autoassociative network can perform pattern completion and retrieve the predicted next state. Levy (1996) presented a model of hippocampal region CA3 as a sequence predictor and argues that this general sequence prediction paradigm can provide a computational unification of a variety of putative hippocampal-dependent functions, including contextual sensitivity, configuration, and cognitive mapping (see also Levy et al 1995, Prepscius & Levy 1994). Granger et al (1996) presented a model of field CA1 incorporating an LTP learning rule in which the amount of potentiation depends on the order of arrival of afferent activity to a target neuron. They show that with this temporally dependent LTP learning, the CA1 network model can learn to store brief simulated temporal sequences of inputs. Liaw & Berger (1996) also described a model of hippocampal neurons in which they argued that the dynamic interplay of hippocampal synaptic mechanisms for facilitative and inhibitory processes results in an emergent "temporal chunking" mechanism for sequential pattern recognition. In this model, each dynamic synapse learns to respond to a small sub-pattern of inputs, and the postsynaptic neuron learns how to properly combine these subpatterns.

SPATIAL MEMORY AND NAVIGATION This aspect of autoassociative memory systems seems ideal for implementing a spatial processor, in which the broad memory of a place should be evoked by any of several views of the area, even if some of the usual cues are missing. In fact, spatial memory is extremely hippocampal dependent in rats (e.g. O'Keefe & Nadel 1978), and many connectionist models of hippocampal-processing in spatial learning have been based on autoassociative models of the hippocampal region (Burgess et al 1994, Levy 1989, McNaughton & Morris 1987, McNaughton & Nadel 1990, Muller et al 1987, Muller & Stead 1996, Recce & Harris 1996, Sharp 1991, Sharp et al 1996). One possibility is to define spatial maps as composed of sets of complex configural associations representing places (McNaughton 1989, McNaughton & Nadel 1990). In one place, there may be many views, depending on which way the animal is facing, the location of landmarks, etc. The hippocampal autoassociator would be able to map from one of these views to the full representation of the current place. With this interpretation, place learning need

not be fundamentally different from any other kind of representational learning. However, because of the need for such complex representations in spatial tasks, these behaviors might be especially sensitive to hippocampal damage.

EPISODIC MEMORY AND CONSOLIDATION Perhaps most pervasive is the idea that the fast, temporary storage in an autoassociator is an important component of an episodic or declarative memory system, in which arbitrary patterns are stored (Alvarez & Squire 1994, Hasselmo et al 1996, McClelland & Goddard 1996, Murre 1996, O'Reilly & McClelland 1994, Treves & Rolls 1992). It is generally assumed in these models that a relatively small temporary store in the hippocampus interacts with a relatively large neocortical system (Figure 4). Such an assumption was made by Marr (1971), and many connectionist models of amnesia center on similar assumptions (e.g. Alvarez & Squire 1994, Lynch & Granger 1992, McClelland et al 1994, Murre 1996, O'Reilly & McClelland 1994, Treves & Rolls 1992). Many preconnectionist models assume this general organization as well (e.g. Mishkin 1982, Teyler & DiScenna 1986, Wickelgren 1979).

In these models, the central assumption is that a stimulus enters the neocortex via the sensory system and subsequently activates cells in the hippocampus. The hippocampus in turn feeds back to the neocortex and initiates activation patterns there. It may activate new cell populations, which are then added to the representation, or it may allow connections to form between

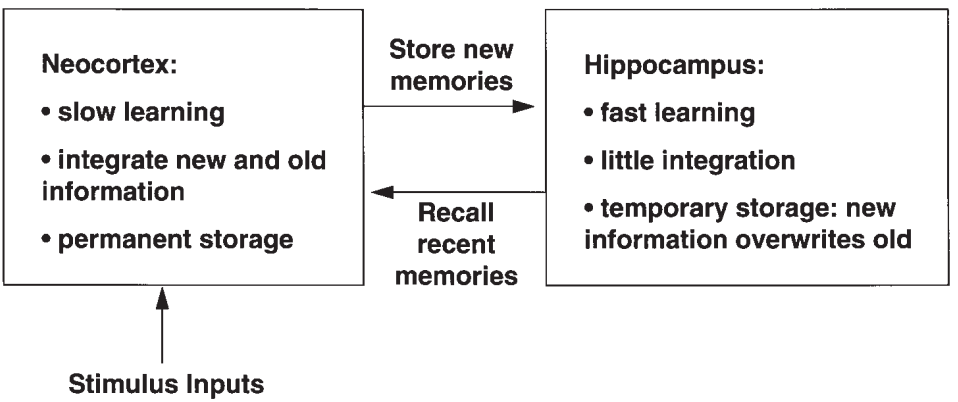

Figure 4 General format of many connectionist models of amnesia. The neocortex is assumed to be a large-capacity, permanent store for memory associations, and to be able to integrate new information with old associations. However, learning is assumed to be slow and possibly require several iterated presentations. The hippocampus is assumed to be capable of storing memory within as little as a single exposure, but older memories are liable to be overwritten by newer ones. The hippocampus therefore captures episodic memories and iteratively allows the neocortex to integrate these memories with existing associations.

active cells in the neocortex. The hippocampus may be required to present memories to the neocortex repeatedly, over some period, to allow the neocortex to integrate new knowledge without overwriting the old (McClelland et al 1994). This process is termed memory *consolidation.* Over time, as this consolidation occurs, the sensory input is able to activate these cells directly, without hippocampal intervention. At this point, the hippocampus has completed its function of helping to bind together disparate cortical activities into a coherent pattern, and memories are safe from subsequent hippocampal damage. However, a more recent memory, which is not yet fully consolidated, may be disrupted. The probability of such disruption is higher for more recent memories, which have had less time to be consolidated, than older ones. This is consistent with data showing that while hippocampal damage leads to severe anterograde amnesia, there is only temporally gradated retrograde amnesia (Squire & Alvarez 1995). This inverse relationship between memory age and hippocampal independence is known as the *Ribot gradient* of retrograde amnesia (Ribot 1882; see also Alvarez & Squire 1994). Examples from animal and human experiments are shown in Figures 5*A* and 5*B* (Kim & Fanselow 1992, Squire & Cohen 1979).

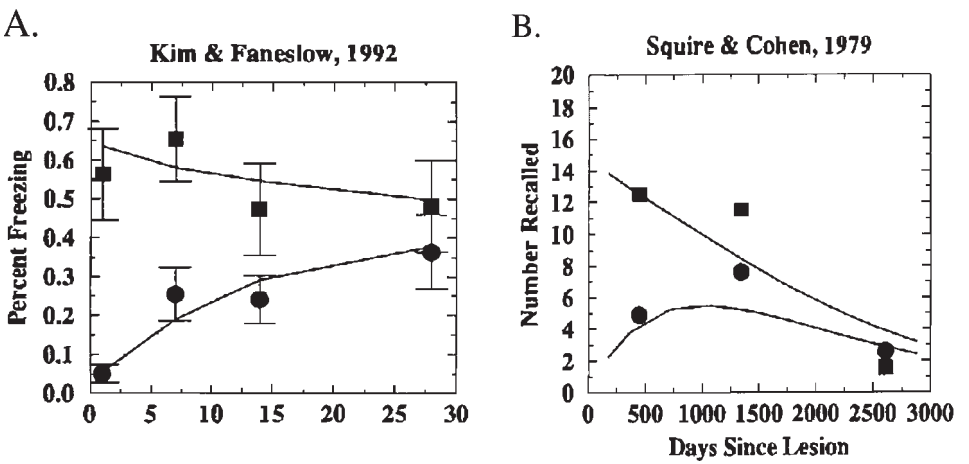

Figure 5 Examples of Ribot gradients, which illustrate how older memories are less likely to be disrupted by hippocampal damage than are newer memories. (*A*) Ribot gradient in animal data. Behavioral responses of animals receiving extensive hippocampal system lesions (*circles*) or control lesions (*squares*) as a function of the number of days elapsing between exposure to the relevant experiences and the occurrence of the lesion. Fear response (freezing) behavior shown by rats when returned to an environment in which they had experienced paired presentations of tones with foot shock. Bars surrounding each datapoint indicate the standard error (from Kim & Faneslow 1992). (*B*) Ribot gradient in human data. Recall by depressed human subjects of details of television shows aired different numbers of years before the time of test, after electroconvulsive treatment (*circles*) or just before treatment (*squares*) (from Squire & Cohen 1979).

Note that a model consisting of an autoassociator alone would predict the opposite effect: namely, that older memories would be increasingly susceptible to interference from newer memories. The addition of a "remote" neocortical storage site allows the models of hippocampal-cortical interaction to account for both the anterograde and retrograde aspects of hippocampal amnesia. Further elaborations may be assumed on this general model scheme, such as nonspecific modulatory influences that determine the storage rates in CA3 (Grossberg 1976, Hasselmo et al 1995, Murre 1996, Treves & Rolls 1992), or additional preprocessing in dentate gyrus and postprocessing in CA1 (Hasselmo & Schnell 1994, Levy 1989, McNaughton 1991, Treves & Rolls 1992).

Open Issue: How and When Does Consolidation Take Place?

A major challenge confronting these models of anterograde amnesia is to specify in detail just how consolidation of memories from hippocampus to neocortex might take place. One small-scale implementation is provided by Alvarez & Squire (1994), who suggested that most memory consolidation may occur during sleep (see also Buzsaki 1989, Crick & Mitchison 1983, McClelland et al 1994). This is consistent with recent data showing that hippocampal activity during slow-wave sleep echoes specific patterns recorded earlier while the animal was exploring its environment (Wilson & McNaughton 1994). Alvarez & Squire suggested that this activity reflects a process during which the hippocampus reinstates patterns it stored earlier and presents them to neocortex for consolidation. The electrical activity in the hippocampus is markedly different during waking exploration and slow-wave sleep, which further suggests that the hippocampus is operating in two different modes (information storage and information reinstatement) during these two behavioral states (Buzsaki 1989). Other possible mechanisms of consolidation may include conscious and unconscious rehearsal (Murre 1996). All these hypotheses await thorough verification through combined neurophysiological and neuropsychological studies.

Open Issue: The Problem of Interference in Memory Networks

Another issue concerns the problem of interference. One constraint on the utility of an autoassociative network is that it has very limited capacity. A network of n nodes is able to store only about $0.15n$ random patterns before they begin to interfere with one another (Hopfield 1982). Interference refers to the likelihood that patterns overlap sufficiently such that retrieval of one will activate retrieval of part or all of additional patterns, and the resulting network output will contain elements of multiple stored patterns. In the extreme, in a net that is filled to capacity, addition of a single new pattern can disrupt the

ability to correctly retrieve any previously stored pattern. As described above, this phenomenon is called catastrophic interference (Hetherington 1990), and it is a general feature of all connectionist networks that perform fast storage, as the hippocampal auto-associative network is assumed to do (McClelland et al 1994). One way to increase capacity and avoid catastrophic interference is to explicitly decrease the overlap between patterns. It has been suggested that this is one effect of the sparse connections from dentate gyrus to CA3: Since any one mossy fiber contacts only about 14 of the 3×10^5 CA3 cells in rat, there is very little probability that two patterns of mossy fiber activity will activate the same pattern of CA3 activity (Rolls 1989). In addition, plasticity in the dentate gyrus may further help to sparsify CA3 inputs (Hasselmo 1995, O'Reilly & McClelland 1994, Treves & Rolls 1992). Even with such *pattern separation*, a pattern stored in the hippocampus will only remain intact for a limited period before it is overwritten by storage of newer memories. This implies that memories stored in the hippocampus must be transferred elsewhere to survive for long periods.

STIMULUS REPRESENTATION IN ASSOCIATIVE LEARNING

The models described above focus on the ability of the hippocampal region to perform fast, temporary storage, and they suggest that this underlies the hippocampal region's role in episodic memory formation. This is consistent with the basic idea that episodic memory impairments are the most obvious behavioral effects in human amnesia following hippocampal region damage. Nondeclarative learning (including procedural or implicit learning) often survives such damage. For example, animals with hippocampal-region damage can often show normal acquisition of classically conditioned responding (e.g. Solomon & Moore 1975) or discrimination of successively presented odors (Eichenbaum et al 1988). Similarly, human hippocampal-damaged amnesics are not impaired at acquiring conditioned motor reflex responses (Daum et al 1989, Gabrieli et al 1995, Woodruff-Pak 1993), learning simple classification tasks (Knowlton et al 1994), or learning new motor skills such as mirror drawing (Cohen 1984). All these tasks can be solved by incremental formation of habits or tendencies, without requiring episodic memories of any individual learning session.

However, there are other tasks that seem superficially to be just as nondeclarative but that are impaired after hippocampal-region damage. For example, although the simplest acquisition of a classically conditioned response is not impaired by hippocampal-region damage, there may be severe impairments in classical conditioning tasks that require learning about unreinforced stimuli

(Kaye & Pearce 1987, Solomon & Moore 1975), configurations of stimuli (Rudy & Sutherland 1989), contextual information (Hirsh 1974), or relationships that span short delays (Moyer et al 1990, Port et al 1986). These findings imply that the hippocampal region does participate in information processing during procedural tasks, although this participation may not necessarily be evident in the simplest kinds of learning. These findings also indicate that a conception of the hippocampal region as a purely passive store for episodic memories is insufficient.

Several recent qualitative theories and computational models have focused on possible information processing roles for the hippocampal region, especially in incrementally acquired (nondeclarative) learning (e.g. Eichenbaum et al 1992b; Gluck & Myers 1993, 1995; Hirsh 1974; Moore & Stickney 1980; Myers et al 1995, 1996; Schmajuk & DiCarlo 1992; Sutherland & Rudy 1989). In turn, these models are less concerned with the issues that motivate the above-described models of consolidation. A full account of hippocampal-region function would, of course, address its role in both information processing and the consolidation of declarative memories.

Most of these associative theories of incremental learning assume that while the hippocampus is required for some complicated forms of stimulus association, the neocortex is sufficient for simpler stimulus-response associations such as those that underlie classical conditioning (e.g. Gluck & Myers 1993; Myers et al 1995; Schmajuk & DiCarlo 1990, 1992). Here we focus on one representative computational model, which incorporates some of the earlier ideas regarding hippocampal autoassociation (Gluck & Myers 1993).

Hippocampal Function and Stimulus Representations

Gluck & Myers (1993) presented a computational theory of hippocampal-region function in associative learning, which argued that the hippocampal region is critical during learning for recoding neural representation to reflect environmental regularities. Central to this theory is the definition of a stimulus representation as a pattern of activities over a set of elements (neuron groups in a brain or nodes in a connectionist network) evoked by the stimulus. Learning to make a response to that stimulus involves mapping from that representation to appropriate behavioral outputs. Learning about one stimulus will transfer—or generalize—to other stimuli as a function of how similar their representations are. Therefore, the particular representations can have a great impact on how hard a task is to learn.

The key idea of Gluck and Myers's (1993) cortico-hippocampal model is that the hippocampal region is able to facilitate learning by adapting representations in two ways. First, it is assumed to compress, or make more similar, representations of stimuli that co-occur; second, it is assumed to differentiate,

or make less similar, representations of stimuli that are to be mapped to different responses. This kind of function can be implemented in a connectionist model that is related to the autoassociators described above but that includes a middle (often termed hidden) layer of nodes. Such a network, termed an autoencoder (Hinton 1989), is shown on the left in Figure 6. It maps input activations representing stimulus inputs through weighted connections to activate the middle layer of nodes that in turn feed through weighted connections to activate the output layer of nodes. The network is trained to produce outputs that reconstruct the inputs as well as predict the behavioral response. Because the autoencoder has a narrow hidden layer of nodes, this task can only be accomplished by compressing redundant information, while preserving and differentiating enough predictive information to allow reconstruction at the output layer. Although the details of the model are not biologically realistic (especially the use of backpropagation error-correction for updating the autoencoder weights), the model nevertheless is a useful tool for exploring the

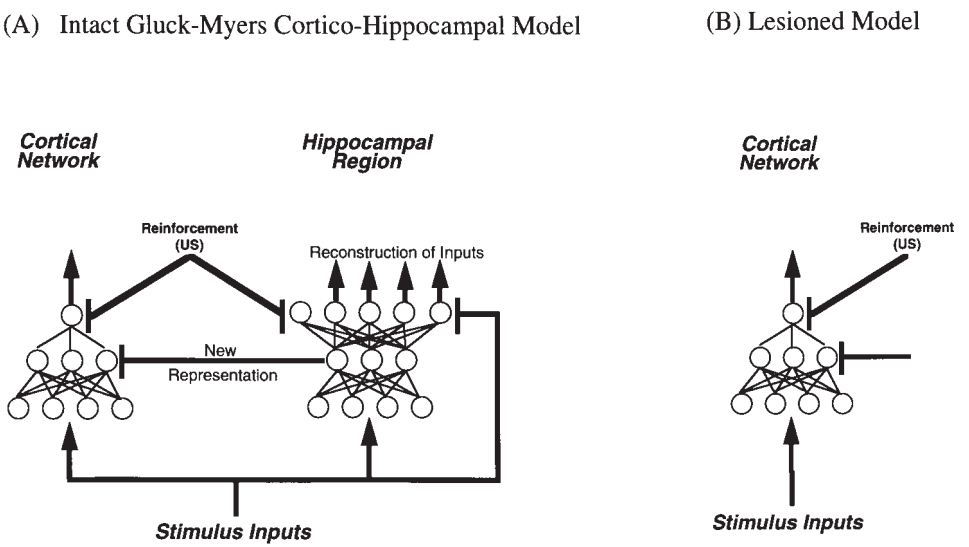

Figure 6 The cortico-hippocampal model (Gluck & Myers 1993). (*A*) The intact system is assumed to include a predictive autoencoder, representing hippocampal-region processing that constructs new stimulus representations in its internal layer that are biased to compress redundancies while differentiating predictive information. These stimulus representations are acquired by long-term storage sites in the cortex, represented as a multilayer network that learns to predict US arrival. The cortical network uses the Rescorla-Wagner rule to map from inputs to the hippocampal-mediated internal representations, and again to map from the internal layer to output activations. (*B*) Hippocampal-region lesion is assumed to disable the hippocampal network, in which case the cortical network can no longer acquire new internal representations but can acquire new behavioral responses based on its preexisting (and now fixed) internal representations. (Reprinted from Myers & Gluck 1995.)

kinds of representations that might evolve under the constraints of the two biases described above (for a more biological instantiation of these same ideas, see section on "Dissociating Parahippocampal and Hippocampal Contributions" and Myers et al 1995).

This network is incorporated into the full cortico-hippocampal model shown in Figure 6 (Gluck & Myers 1993). A cortical network is shown on the left, which is assumed to map from stimulus inputs to outputs that determine the behavioral response. However, this network is assumed to be unable to construct hidden layer representations on its own. Instead, it can adopt those representations formed in the hidden layer of the hippocampal region network. It can then learn to map from these to the correct responses. Hippocampal lesion is simulated in this model by disabling the hippocampal network, and assuming that the hidden layer representations in the cortical network are now fixed. Those already acquired are maintained, so little retrograde amnesia is expected after hippocampal-region damage (although the model does not rule out the idea of an indefinitely long consolidation period during which information is transferred, as suggested by the models of McClelland, Murre, and others reviewed above). Further, the cortical network can still learn to map from the existing representations to new behavioral responses. All that is lost is the hippocampal-dependent ability to modify those representations.

Application to Behavioral Data

Gluck & Myers's (1993) model can be applied to classical conditioning by assuming that the inputs are conditioned stimuli, and that the output is a conditioned response that is expected to anticipate the reinforcing unconditioned stimulus. The model then captures many aspects of the behavior of intact and hippocampal-region–damaged animals (Gluck & Myers 1993, 1996; Myers & Gluck 1994, 1996). For example, the model correctly expects that hippocampal-region damage causes no particular impairment—or even a slight facilitation—in learning a conditioned response. For such a simple task, new adaptive representations are probably not needed, and even the lesioned model can learn to map from its existing representations to the correct response. In fact, because the intact model is slowed by constructing new representations, it may often be slower than the lesioned model. This is consistent with similar effects often seen in animals (e.g. Eichenbaum et al 1988, Schmaltz & Theos 1972, etc).

However, latent inhibition—the slower learning after unreinforced exposure to the to-be-conditioned stimulus (Lubow 1973)—is disrupted by broad hippocampal-region damage (Figure 7*A*) (Kaye & Pearce 1987, Solomon & Moore 1975). The model correctly shows these effects (Myers & Gluck 1994). During the exposure phase, the stimulus is partially redundant with the back-

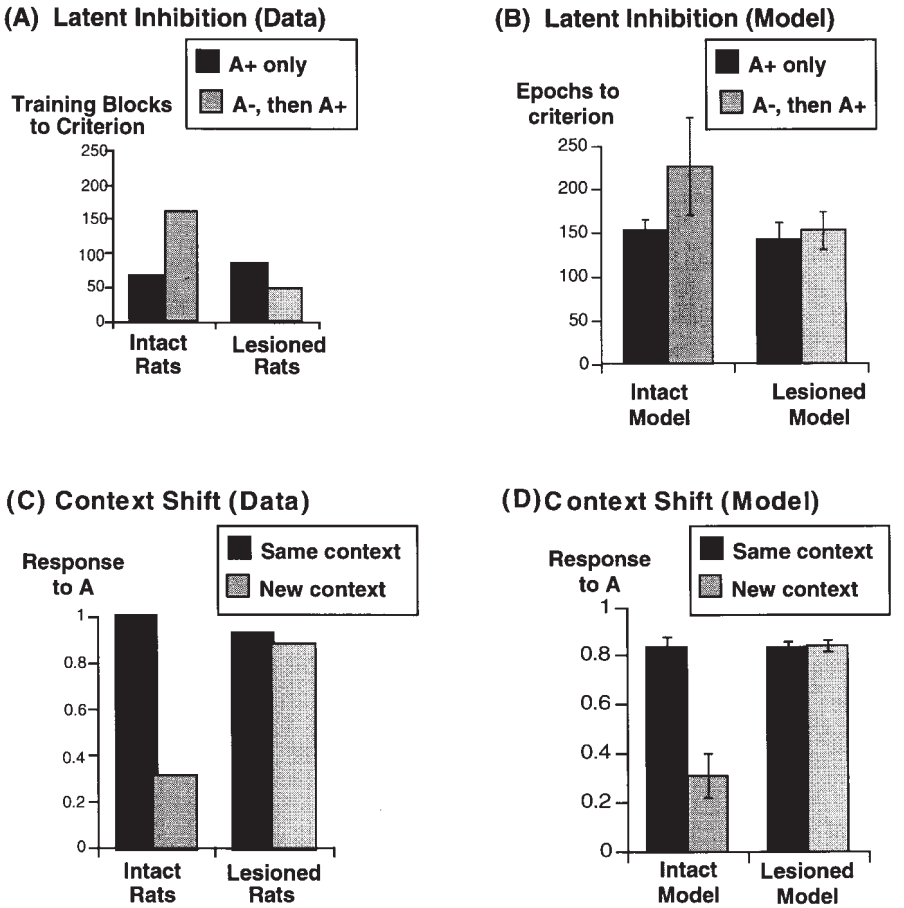

Figure 7 Behavioral results from intact and hippocampal-lesioned animals compared with simulation results from intact and lesioned cortico-hippocampal model. (*A*) Latent inhibition. In intact animals, unreinforced preexposure to a cue A slows later acquisition of conditioned responding to A (Lubow 1973). This is reflected in longer training times until criterion is reached on responding to A. Broad hippocampal-region lesion eliminates this effect (Kaye & Pearce 1987, Solomon & Moore 1975). (Figure plotted from data presented in Solomon & Moore 1975.) (*B*) The intact model correctly shows latent inhibition, whereas the lesioned model does not. (Figure reprinted from Myers et al 1995.) (*C*) In normal animals, a conditioned response to A may show a decrement if A is then presented in a new context (Hall & Honey 1989). Hippocampal-lesioned animals do not show this response decrement after a context shift (Honey & Good 1993, Penick & Solomon 1991). (Figure replotted from data presented in Penick & Solomon 1991.) (*D*) The intact but not lesioned model correctly shows this response decrement with context shift (Myers & Gluck 1994). (Figure reprinted from Myers et al 1995.)

ground context. Neither predicts any reinforcing event, so the hippocampal-region network compresses their representations. Later, when the task is to respond to the stimulus but not the context alone, this compression must be undone, which results in slowed learning in the intact model (Figure 7*B*). In contrast, the lesioned model has no compression during the exposure phase, so learning is not retarded in the subsequent learning phase (Figure 7*B*).

Many of the learning deficits associated with hippocampal damage can be described as context effects (Hirsh 1974). For example, human hippocampal-damaged amnesics may be able to remember an experience but not where or when that information was acquired—and they may even be unaware they know the information itself until indirectly prompted for it (Haist et al 1991, Weiskrantz & Warrington 1979). Animals show related effects. For example, under some conditions, an animal trained to respond to a stimulus in one environment gives a decremented response when that stimulus is presented in another environment (Figure 7*C*) (Hall & Honey 1989). A hippocampal-lesioned animal does not show this decrement but responds just as strongly in the new environment (Honey & Good 1993, Penick & Solomon 1991). The cortico-hippocampal model implies a similar effect (Figure 7*D*) (Myers & Gluck 1994) because the hippocampal-region autoencoder is assumed to reconstruct not only the conditioned stimuli but also any background or context cues that are present during learning. Thus, as the autoencoder learns to represent a conditioned stimulus, information about the context is included in that representation. As a result, if the stimulus is presented in a new context, the representation of that stimulus will be less weakly activated than usual, and in turn the conditioned response will be decremented, just as observed in intact animals. In contrast, the lesioned model does not form new, compressed representations, and so responding does not drop in a new context.

In the same way that the cortico-hippocampal model can account for latent inhibition and context shift phenomena, it can similarly address results from a range of conditioning studies (Gluck & Myers 1993, 1996; Myers & Gluck 1994, 1996). It provides a computational instantiation of several prior qualitative theories that posited hippocampal region roles in context learning (Hirsh 1974), configural learning (Sutherland & Rudy 1989), and representational learning (Eichenbaum & Bunsey 1995).

Open Issues and Alternative Approaches

The most obvious limitation of the cortico-hippocampal model, like others in the same domain, is that it does not make any particular attempt to address the episodic memory deficits that are the most obvious feature of human hippocampal amnesia. This is the converse of the limitation of models that address episodic memory but not information processing in the hippocampal region.

Eventually, a complete model of hippocampal-region function will have to account for both these aspects of hippocampal-region damage. For now, though, these models should be judged on the basis of how well they account for the circumscribed set of data they attempt to address.

There are also several limitations of Gluck & Myers's cortico-hippocampal model. As a trial-level model, it cannot capture any of the intricacies of timing within a trial—such as the effects of varying stimulus scheduling, the latency of onset of the conditioned response, and so on. Other models (e.g. Schmajuk & DiCarlo 1990, 1992) do include real-time effects in their models, and they capture these aspects of animal learning. In the next section we consider in more detail another model of Schmajuk and colleagues that addresses a similar body of behavioral conditioning data.

A more general limitation of this entire class of models is the restricted degree of physiological realism they involve. The network architectures and learning algorithms are determined more by functional (behavioral) considerations than by biological properties. In fact, most of these models include properties that are clearly unrealistic, e.g. full or near-full connectivity between sets of nodes. Some attempts have been made to address this limitation. These are reviewed in the next section, which considers more recent attempts to take abstract theories of hippocampal region function and clarify more precisely the functional role of the different anatomical components of this region.

DISSOCIATING PARAHIPPOCAMPAL AND HIPPOCAMPAL CONTRIBUTIONS

Recent refinements in lesion techniques indicate that the extent of memory impairment often depends critically on exact lesion extent. This suggests that the different substructures of the hippocampal region have differentiable contributions to the processing of the region as a whole. However, the precise assignment of function to substructure, and the ways in which they interact, are as yet poorly understood. One example is the latent inhibition effect described earlier, in which prior unreinforced exposure to a stimulus retards later learning to respond to that stimulus (Lubow 1973). Latent inhibition is attenuated by broad hippocampal-region damage (Kaye & Pearce 1987, Solomon & Moore 1975) but not by damage strictly limited to the hippocampus and sparing entorhinal cortex (Honey & Good 1993, Reilly et al 1993). Similarly, odor discrimination reversal is impaired by hippocampal lesion but actually facilitated after entorhinal lesion (Otto et al 1991).

Although the representational theory of hippocampal function proposed by Gluck & Myers (1993) treated the hippocampal region as a single processing

system, subsequent work by these researchers have suggested how their basic representational processes might be subdivided, and the subfunctions localized in different anatomical sites around the region (Myers et al 1995). In particular, Myers et al (1995) proposed that stimulus-stimulus redundancy compression could emerge from the anatomy and physiology of superficial entorhinal cortex.

Parahippocampal Function in Stimulus Compression and Clustering

The Myers et al model of entorhinal (and parahippocampal) function in learning is derived from an earlier physiologically realistic model of superficial piriform (olfactory) cortex by Ambros-Ingerson et al (1990), which argued that the anatomy and physiology of this cortical structure are sufficient to implement hierarchical clustering of odor inputs. In brief, Ambros-Ingerson et al proposed a competitive network model in which local recurrent inhibition silences all but the most strongly responding pyramidal cells. These so-called winning cells come to respond to a family or cluster of inputs with similar features. Recurrent feedback from the piriform cortex to olfactory bulb is also assumed to allow iterative responses to odors, from which successively finer-grained (hierarchical) classifications can be constructed. One aspect of this model is that, since similar inputs tend to be clustered to similar output responses, the network performs redundancy compression of exactly the sort previously proposed by Gluck & Myers (1993) to occur in the hippocampal region (Myers et al 1995). In particular, if two inputs co-occur, they will be treated as a single compound input. Later, if one of the inputs occurs alone, the network will tend to treat this as a degraded version of the compound input and assign it to the same cluster as the compound.

The piriform cortex and entorhinal cortex elide in rat, and their superficial layers are closely related anatomically and physiologically, suggesting the possibility of related functionality (Price 1973, van Hoesen & Pandya 1975, Woodhams et al 1993). Specifically, superficial entorhinal cortex contains pyramidal cells with sparse nontopographic connections with afferents in layer I (van Hoesen & Pandya 1975) with denser feedback connections to local inhibitory cells (Kohler 1986), and shows NMDA-dependent, theta-induced long-term potentiation (LTP) (de Curtis & Llinas 1993). Noting this similarity, Gluck & Granger (1993) suggested that entorhinal cortex could perform a similarity-based clustering operation similar to that proposed to occur in piriform cortex.

In sum, then, Myers et al (1995) have proposed that the entorhinal cortex would be sufficient to implement the redundancy compression aspect of the representational changes that Gluck & Myers (1993) ascribe to the hippocampal region as a whole (Myers et al 1995). A model implementing these proposed processes, and based on the physiologically and anatomically motivated

model of Ambros-Ingerson et al (1990), is shown in Figure 8*A*. One difference between the piriform and entorhinal models is that the piriform model assumes repetitive sampling and input masking, based on recurrent connections from piriform cortex to olfactory bulb. Myers et al (1995) have not assumed this in the entorhinal model, and so it only performs a single-stage, similarity-based clustering or compression of its inputs. The resulting network is similar to the unsupervised, competitive-learning systems developed by Kohonen (1984), Rumelhart & Zipser (1985), Grossberg (1976), and others. A second important difference between the piriform and entorhinal cortices is that, while the piriform cortex is primarily an olfactory area, the entorhinal cortex receives input from a broad spectrum of polymodal cortices, as well as from the piriform cortex. Thus, Myers et al (1995) have suggested that while the piriform cortex might be sufficient to implement redundancy compression within the olfactory domain, the entorhinal cortex might be required to implement redundancy compression between stimuli from different modalities, or across the polymodal features of a single stimulus (Myers et al 1995).

This model can be compared with a lesion that selectively damages the hippocampus and dentate gyrus but that leaves intact the entorhinal cortex. As noted above, such lesions often produce different results from lesions of the entire hippocampal region. For example, such a restricted lesion does not disrupt latent inhibition, although as described above a larger lesion does (Honey & Good 1993, Reilly et al 1993). The selectively lesioned model produces the same effect (Figure 8*B*). The redundancy compression in the entorhinal network is sufficient to mediate latent inhibition. The model accounts for several other selective-lesion effects (Myers et al 1995), as well as makes specific novel predictions that other behaviors, which are interpreted as reflecting stimulus compression, are likely to depend more on the entorhinal cortex than on hippocampus proper, and so should survive such a localized lesion.

The idea that the entorhinal cortex is involved in stimulus compression also relates to a suggestion by Eichenbaum & Bunsey (1995) that the entorhinal cortex performs "fusion" of coincident or nearly coincident stimuli, based on the tendency of animals with selective hippocampal (but not entorhinal) damage to overcompress stimulus information.

This hypothesis regarding the selective contribution of entorhinal processing to hippocampal-region function assumes that the remaining subfunction of predictive differentiation could be implemented elsewhere in the hippocampal region. One possibility is that the dentate gyrus or hippocampus proper could perform this subfunction. This idea is consistent with several suggestions that the hippocampus is involved in predicting future events (such as US arrival) given current inputs (e.g. Gray 1985, Levy 1985, Lynch & Granger 1992, McNaughton & Nadel 1990, Treves & Rolls 1992).

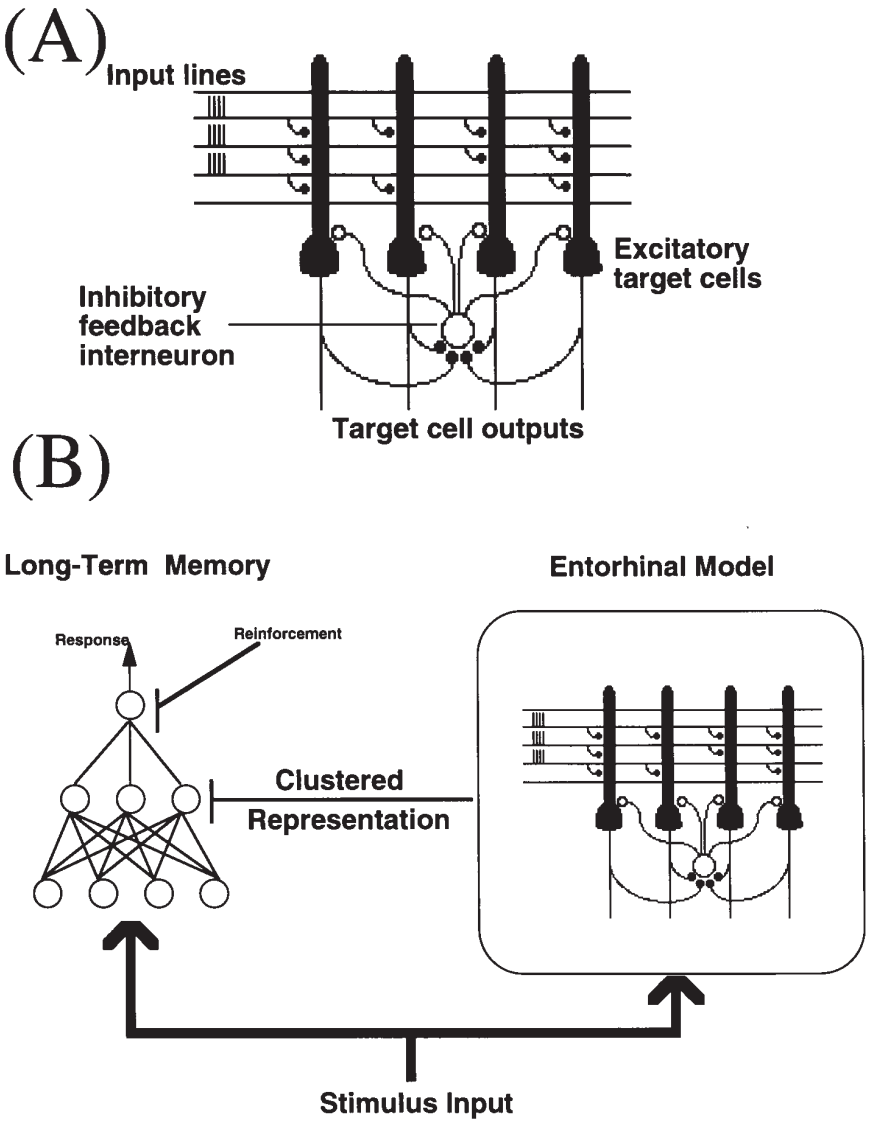

Figure 8 (*A*) In the entorhinal model, target cells are excited by sparse afferents, and in turn activate local inhibitory feedback interneurons. Feedback silences all but the most strongly activated target cells. Synaptic plasticity makes these "winning" target cells more likely to "win" in response to similar inputs in the future. The resulting network activity is constrained by stimulus-stimulus redundancy compression. (*B*) The H-lesioned model, in which an entorhinal cortex network provides new compressed representations to the internal layer of a long-term memory network. (Adapted from Myers et al 1995.)

Parahippocampal Function in Configural Associations

In an alternative approach to modeling entorhinal function, Schmajuk and Blair (Schmajuk 1994, Schmajuk & Blair 1993) have suggested the particular contribution of the entorhinal cortex to the Schmajuk-DiCarlo (Schmajuk & DiCarlo 1992) model of hippocampal-region function is stimulus competition, while the hippocampus proper is responsible for configural association. They therefore predict that localized hippocampal lesion, which does not otherwise damage entorhinal cortex, should eliminate the configural but not the stimulus competition function. Empirical data are somewhat consistent with this idea (see Schmajuk 1994), although further empirical studies are certainly indicated, as mentioned above in the context of testing our own model of entorhinal function. The stimulus competition function proposed by Schmajuk & Blair is quite distinct from the stimulus-stimulus clustering we have proposed as an entorhinal function. In fact, our entorhinal stimulus-stimulus clustering is probably more closely related to the configural function that Schmajuk & Blair assign not to the entorhinal cortex but to the hippocampus proper. Until such time as more empirical data become available, it may be difficult to provide a definitive discrimination between these two accounts. However, future experiments that address the selective role of the entorhinal cortex in stimulus competition and in stimulus-stimulus clustering are required to properly evaluate these two models.

In a more recent paper, Buhusi & Schmajuk (1996) presented a different model of hippocampal function in conditioning that attributes both attentional and configural mechanisms to specific components of the hippocampal region. Buhusi & Schmajuk proposed that the entorhinal and parahippocampal cortices have a unique role in error-correction in which expected reinforcement is compared with actual reinforcement. In contrast, we have argued that these same overlying cortices are essential for stimulus-stimulus redundancy compression. This is consistent with studies showing that latent inhibition, a result Myers et al (1995) have interpreted as being mediated by stimulus compression, is spared after hippocampal lesions that do not extend to entorhinal cortex (Honey & Good 1993, Reilly et al 1993).

INCORPORATING SUBCORTICAL CHOLINERGIC MODULATION

The models of episodic memory and consolidation reviewed in the section on "Autoassociative Models of CA3 and Episodic Memory" are fairly abstract in that there is no particular mapping of nodes and connections to neurons and synapses. As Hasselmo and colleagues have shown, however, it is possible to construct autoassociative models that are much more physiologically realistic.

In this vein, Hasselmo & Schnell (1994; see also Hasselmo et al 1995) have developed a model of laminar connections in the hippocampus to study the possible function of the strong cholinergic input from the medial septum. These authors have suggested that the function of this cholinergic input is to allow the hippocampus to switch between pattern storage and pattern retrieval states. When a new pattern is presented to an autoassociative network as a forcing input, it will activate some of the nodes in the network. Activation from these nodes will travel through the recurrent feedback connections to activate other nodes, and after several iterations, this runaway excitation may result in the entire network becoming active, rather than just those nodes associated with the pattern to be stored. To avoid this runaway excitation, an autoassociative network is usually assumed to operate in two modes, a storage mode during which forcing inputs are present but feedback collaterals are suppressed, and a recall mode, during which there is no forcing input, and recurrent collaterals are allowed to activate nodes. In the context of a network model, it is easy to define two such disparate states.

If hippocampal field CA3 is assumed to operate as an autoassociative network, with mossy fiber afferents providing the forcing inputs, there must be a physiological mechanism to suppress activity on the recurrent collaterals during storage. Hasselmo (Hasselmo 1995, Hasselmo & Schnell 1994) proposed that the septal cholinergic input can provide this switch. Briefly, he suggested that cholinergic input suppresses the recurrent collaterals to allow storage of the new pattern without runaway excitation. When cholinergic input is absent, and entorhinal inputs activate a few CA3 cells, feedback connections recruit more cells to activity, until a stored pattern is recalled and instated on the CA3 nodes. Hasselmo (Hasselmo 1995, Hasselmo & Schnell 1994) further proposed a scheme whereby CA3 can self-regulate this cholinergic input, allowing the hippocampus to recognize when a new pattern should be stored, and signal the septum to send the cholinergic input that allows storage to proceed. In model simulations, such self-regulated suppression of recurrent collaterals does suffice to allow switching between storage and recall states in an autoassociative network (Hasselmo & Schnell 1994). In empirical support of this hypothesis, Hasselmo et al (1995) have shown that a cholinergic agonist carbachol does suppress activity of CA3 cells in slice more in the stratum radiatum, where the recurrent collaterals afferent CA3 dendrites, than in the stratum lucidum, where the mossy fibers afferent CA3 dendrites. Further support comes from findings of anterograde amnesia after medial septal lesion (Berry & Thompson 1979) or pharmacological disruption through anticholinergic drugs such as scopolamine (Solomon et al 1983), consistent with Hasselmo's prediction that cholinergic input is necessary for storage of new information in the hippocampus (Hasselmo 1995, Hasselmo & Schnell 1994).

In the next subsection, we discuss how this cholinergic model of Hasselmo can be related to independently developed models of hippocampal function in classical conditioning (Myers et al 1996) reviewed above.

Septohippocampal Cholinergic Modulation in Conditioning

Myers and Gluck, in collaboration with Hasselmo and Solomon, have recently shown how a simplified version of Hasselmo's cholinergic hypothesis can be instantiated within the Gluck & Myers model, to provide an interpretation of Solomon's data on the behavioral consequences of anticholinergic drugs on classical conditioning. In brief, the integrated model assumes that the tendency of the hippocampal-region network to store new information, as opposed to simply processing it and recalling old information, is determined by the hippocampal-region network's learning rate (Myers et al 1996). Disrupting septal input can therefore be approximated within the Gluck & Myers model by lowering this learning rate—although not the rate at which this information is transferred to the cortical network, nor the rate at which cortical associations develop. The consequence of this depressed hippocampal learning rate is to prolong the initial nonresponding phase before onset of the initial conditioned responses (Figure 8*B*), much as is seen in the experimental data (Figure 8*A*). This computational model of cholinergic function in conditioning is broadly consistent with an earlier suggestion by Thompson & Berry (1979), who argued that the medial-septum is involved primarily in early attentional stages of learning rather than subsequent associational processes.

With this interpretation of cholinergic function, Myers et al (1996) showed that the Gluck & Myers model correctly expects that hippocampal disruption retards conditioning even though outright hippocampal lesion does not. This apparent paradox has previously been noted in the animal literature (Solomon et al 1983), and the model provides insight into why it might be so. Further, the model predicts that if lowering hippocampal learning rates retards learning, increasing learning rates may speed it (Myers et al 1996). This is consistent with data showing that cholinergic agonists can improve learning in subjects with abnormally reduced levels of brain acetylcholine (for a review, see Myers et al 1996). However, in the model, increasing hippocampal learning rates beyond some optimal level actually results in degraded learning, as the network becomes unstable (Myers et al 1996). Therefore, the model predicts that cholinergic therapy should only be transiently effective in normal subjects. In fact, this is the case: While cholinergic agonists at moderate doses tend to improve learning, higher doses may either result in no facilitation or actually impair learning (for a review, see Myers et al 1996). The model therefore provides an account for this empirical phenomenon, which has been problematic in the clinical pharmacology literature.

An alternative approach to modeling septohippocampal cholinergic pathways is the model of Buhusi & Schmajuk (1996). These authors interpret the septohippocampal cholinergic pathways as providing an error-signal that drives learning. In contrast, Myers et al (1996) argued that these pathways can be functionally interpreted as providing modulation of learning rates, which builds upon similar arguments by Hasselmo (see Hasselmo et al 1996). Despite different functional interpretations of the medial septal inputs, both the Buhusi & Schmajuk (1996) and the Myers et al (1996) models correctly expect that cholinergic antagonists (such as scopolamine) should impair acquisition, but not latent inhibition. Buhusi & Schmajuk have not, however, addressed the detailed aspects of learning curves that are analyzed by Myers et al (1996).

SUMMARY AND GENERAL DISCUSSION

We have reviewed several computational models of hippocampal function in learning and memory, concentrating on those that make strongest contact with psychological issues and data from behavioral experiments. Many of these models can be traced to the influential early model of Marr (1971) that, in turn, built upon Hebb's (1949) ideas on how associations are acquired between groups of cell assemblies in the brain. The basic network architecture described by Marr's theory is known as an autoassociator that learns to associate all components of an input pattern with all other components of the same pattern.

Many subsequent researchers have used Marr's basic framework for modeling episodic or event memories in the hippocampus, especially within hippocampal field CA3 that shares many of the basic connectivity requirements for an autoassociator (Hasselmo et al 1996, McNaughton & Nadel 1990, McNaughton & Morris 1987, Rolls 1989). These models focus on the ability of the hippocampal region to perform fast, temporary storage, which suggests that this underlies the hippocampal region's role in episodic memory formation. This is consistent with the neuropsychological data showing that episodic memory impairments are the most obvious behavioral effects in human amnesia following hippocampal region damage. Variations on the hippocampal autoassociator model have been developed to explain sequential learning (for reviews, see Granger et al 1996, Levy 1996, Liaw & Berger 1996), spatial navigation (Burgess et al 1994, Levy 1989, McNaughton & Morris 1987, McNaughton & Nadel 1990, Muller 1987, Muller & Stead 1996, Recce & Harris 1996, Sharp et al 1996, Sharp 1991), and the consolidation of episodic memories (Alvarez & Squire 1994, McClelland & Goddard 1996, Murre 1996, O'Reilly & McClelland 1994, Treves & Rolls 1992).

Another class of hippocampal models have focused on hippocampal involvement in incrementally learned associative habits, such as classical condi-

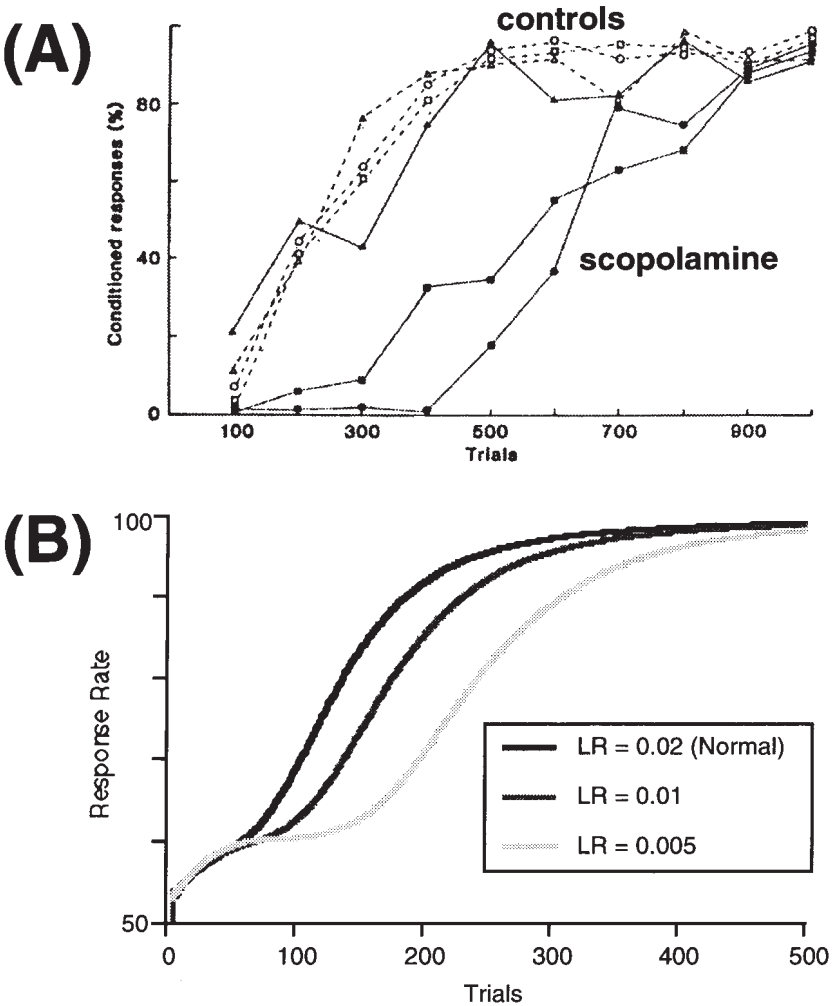

Figure 9 Experimental data and modeling of the effects of the anticholinergic drug scopolamine on acquisition of a conditioned eyeblink response. (*A*) Systemic application of scopolamine (Solomon et al 1983) in which it is shown that the effect of scopolamine is to delay the onset of conditioning, rather than preventing it. (*B*) Learning curves for three different hippocampal learning rates in the Myers et al (1996) model showing how lowered learning rates shift the acquisition curve to the right, delaying the onset of learning, much as seen in Figure 9*A*.

tioning or probabilistic pattern classification. Many recent qualitative theories and several computational models have focused on possible information processing roles for the hippocampal region that are most evident from studying

complex training procedure in incrementally acquired learning (Buhusi & Schmajuk 1996; Eichenbaum et al 1992b; Gluck & Myers 1993, 1995, 1996; Hirsh 1974; Moore & Stickney 1980; Schmajuk & DiCarlo 1992; Sutherland & Rudy 1989). More recent modeling efforts have attempted to make closer contact with the underlying anatomy and physiology. This includes models of parahippocampal function (Myers et al l995, Schmajuk & Blair 1993) and models of the subcortical influences of cholinergic modulation (Buhusi & Schmajuk 1996, Myers et al 1996).

In reviewing these psychobiological models of hippocampal function in learning and memory, three major themes have emerged. First, we have seen how computational models can provide the "glue" to bind together analysis and data at multiple levels of analysis including cellular, physiological, anatomical, and behavioral levels. In particular, we noted how some models are developed in a top-down fashion, beginning with detailed behavioral analyses and then seeking a mapping to underlying biological substrates. Other models are developed in a more bottom-up fashion, beginning with biological details and, via computational simulations, seeking to identify emergent functional properties of these substrates (for further discussion of these distinctions in learning models, see Gluck & Granger 1993).

A second theme that emerged was the importance of models as tools to integrate data from both animal and human studies of hippocampal function in learning and memory. Although these two bodies of research have often been quite separate and disconnected, it seems clear that ultimately they must converge so that each body of literature and theory can inform the other, which will hopefully lead to a more general and broadly applicable understanding of the hippocampal region in all species.

Finally, a third theme that emerged from reviewing these models is the importance of relating current computational models to earlier traditions in memory research, especially the many earlier psychological models that capture important behavioral principles of memory. In drawing these connections between current models, and earlier qualitative theories in psychology and neurobiology, one can see to what extent the models represent cumulative progress.

All the models reviewed here represent preliminary attempts to incorporate both biological data and behavioral analysis within formal computationally defined theories. Crude approximations at best, the value of these models will become most clearly apparent if they lead to important new empirical studies that will inform and constrain future generations of models and theories.

ACKNOWLEDGMENTS

For their helpful comments and suggestions on early drafts of this manuscript, we are grateful to Gyorgy Buzsaki, Michael Hasselmo, Brandon Ermita, Kari

Hoffman, and Todd Allen. This research was supported by grants to MAG from the Office of Naval Research through the Young Investigator program and by grant N00014-88-K-0112, as well as a grant from the McDonnell-Pew Program in Cognitive Neuroscience.

Literature Cited

Alvarez P, Squire L. 1994. Memory consolidation and the medial temporal lobe: a simple network model. *Proc. Natl. Acad. Sci. USA* 91:7041–45

Ambros-Ingerson J, Granger R, Lynch G. 1990. Simulation of paleocortex performs hierarchical clustering. *Science* 247:1344–48

Anderson J. 1977. Neural models with cognitive implications. In *Basic Processes in Reading: Perception and Comprehension*, ed. D LaBerge, S Samuels, pp. 27–90. Hillsdale, NJ: Erlbaum

Berry S, Thompson R. 1979. Medial septal lesions retard classical conditioning of the nictitating membrane response in rabbits. *Science* 205:209–11

Bliss T, Lomo T. 1973. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.* 232:331–56

Buhusi C, Schmajuk N. 1996. Attention, configuration, and hippocampal function. *Hippocampus.* In press

Burgess N, Recce M, O'Keefe J. 1994. A model of hippocampal function. *Neural Networks* 17:1065–81

Buzsaki G. 1989. Two-stage model of memory-trace formation: a role for "noisy" brain states. *Neuroscience* 31(3):551–70

Cohen N. 1984. Preserved learning capacity in amnesia: evidence for multiple learning systems. See Squire & Butters 1984, pp. 83–103

Crick F, Mitchison G. 1983. The function of dream sleep. *Nature* 304:111–14

Daum I, Channon S, Canavan A. 1989. Classical conditioning in patients with severe memory problems. *J. Neurol. Neurosurg. Psychiatry* 52:47–51

de Curtis M, Llinas R. 1993. Entorhinal cortex long-term potentiation evoked by theta-patterned stimulation of associative fibers in the isolated in vitro guinea pig brain. *Brain Res.* 600:327–30

de Leon M, Golomb J, George A, Convit A, Rusinek H, et al. 1993. Hippocampal formation atrophy: prognostic significance for Alzheimer's disease. In *Alzheimer's Disease: Advances in Clinical and Brain Research*, ed. B Corain, K lqbal, M Nicolini, B Winblad, H Wisniewski, P Zatta, pp. 35–46. New York: Wiley

Eichenbaum H. 1992. The hippocampal system and declarative memory in animals. *J. Cogn. Neurosci.* 4(3):217–31

Eichenbaum H, Bunsey M. 1995. On the binding of associations in memory: clues from studies on the role of the hippocampal region in paired associate learning. *Curr. Dir. Psychol. Sci.* 4:19–23

Eichenbaum H, Cohen NJ, Otto T, Wible C. 1992. Memory representation in the hippocampus: functional domain and functional organization. In *Memory Organization and Locus of Change*, ed. LR Squire, G Lynch, NM Weinberger, JL McGaugh, pp. 163–204. Oxford: Oxford Univ. Press

Eichenbaum H, Fagan A, Mathews P, Cohen NJ. 1988. Hippocampal system dysfunction and odor discrimination learning in rats: impairment or facilitation depending on representational demands. *Behav. Neurosci.* 102(3):331–39

Gabrieli J, McGlinchey-Berroth R, Carrillo M, Gluck M, Cermack L, Disterhoft J. 1995. Intact delay-eyeblink classical conditioning in amnesia. *Behav. Neurosci.* 109(5): 819–27

Gluck MA, Granger R. 1993. Computational models of the neural bases of learning and memory. *Annu. Rev. Neurosci.* 16:667–706

Gluck MA, Myers C. 1993. Hippocampal mediation of stimulus representation: a computational theory. *Hippocampus* 3:491–516

Gluck M, Myers C. 1995. Representation and association in memory: a neurocomputational view of hippocampal function. *Curr. Dir. Psychol. Sci.* 4(1):23–29

Gluck MA, Myers CE. 1996. Integrating behavioral and physiological models of hippocampal function. *Hippocampus.* (Special Issue on Hippocampal Computation & Memory; M Gluck, Guest Editor)

Granger R, Wiebe S, Taketani M, Lynch G. 1996. Distinct memory circuits comprising the hippocampal region. *Hippocampus.* In press

Gray JA. 1985. Memory buffer and comparator can share the same circuitry. *Behav. Brain Sci.* 8(3):501

Grossberg S. 1976. Adaptive pattern classification and recoding: part I. *Biol. Cybern.* 23: 121–34

Haist F, Musen G, Squire LR. 1991. Intact priming of words and nonwords in amnesia. *Psychobiology* 19(4):275–85

Hall G, Honey RC. 1989. Contextual effects in conditioning, latent inhibition, and habituation: associative and retrieval functions of contextual cues. *J. Exp. Psychol.: Anim. Behav. Process.* 15(3):232–41

Hasselmo ME. 1995. Neuromodulation and cortical function: modeling the physiological basis of behavior. *Behav. Brain Res.* 67:1–27

Hasselmo ME, Schnell E. 1994. Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CAl: computational modeling and brain slice physiology. *J. Neurosci.* 14: 3898–914

Hasselmo ME, Schnell E, Barkai E. 1995. Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *J. Neurosci.* 15:5249–62

Hasselmo ME, Wyble B, Wallenstein G. 1996. Encoding and retrieval of episodic memories: role of cholinergic and GABAergic modulation in the hippocampus. *Hippocampus.* In press

Hebb DO. 1949. *The Organization of Behavior.* New York: Wiley

Hetherington P. 1990. *The sequential learning problem in connectionist networks.* MS thesis. McGill Univ.

Hinton G. 1989. Connectionist learning procedures. *Artif. Intell.* 40:185–234

Hirsh R. 1974. The hippocampus and contextual retrieval of information from memory: a theory. *Behav. Biol.* 12:421–44

Honey RC, Good M. 1993. Selective hippocampal lesions abolish the contextual specificity of latent inhibition and conditioning. *Behav. Neurosci.* 107(1):23–33

Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* 79:2554–58

Kaye H, Pearce J. 1987. Hippocampal lesions attenuate latent inhibition and the decline of the orienting response in rats. *Q. J. Exp. Psychol.* 39:107–25

Kelso SR, Ganong AH, Brown TH. 1986. Hebbian synapses in hippocampus. *Proc. Natl. Acad. Sci. USA* 83:5326–30

Kim JJ, Fanselow MS. 1992. Modality-specific retrograde amnesia of fear. *Science* 256: 676

Knowlton B, Squire L, Gluck M. 1994. Probabilistic classification learning in amnesia. *Learn. Mem.* 1:106–20

Kohler C. 1986. Intrinsic connections of the retrohippocampal region in the rat brain. II. The medial entorhinal area. *J. Comp. Neurol.* 246:149–69

Kohonen T. 1984. *Self-Organization and Associative Memory.* New York: Springer-Verlag

Levy W. 1985. An information/computation theory of hippocampal function. *Soc. Neurosci. Abstr.* 11:493

Levy WB. 1989. A computational approach to hippocampal function. In *Psychology of Learning and Motivation,* R Hawkins, G Bower, pp. 243–304. London: Academic

Levy WB. 1996. A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus.* In press

Levy WB, Brassel SE, Moore SD. 1983. Partial quantification of the associative synaptic learning rule of the dentate gyrus. *Neuroscience* 8(4):799–808

Levy W, Wu X, Baxter R. 1995. Unifications of hippocampal function via computational/encoding considerations. *Int. J. Neur. Syst.* 6(Suppl.): 71–80

Liaw J, Berger TW. 1996. Dynamic synapses: dancing to the neural tune. *Hippocampus.* In press

Lubow R. 1973. Latent inhibition. *Psychol. Bull.* 79:398–407

Lynch G, Granger R. 1992. Variations in synaptic plasticity and types of memory in corticohippocampal networks. *J. Cogn. Neurosci.* 4(3):189–99

Marr D. 1971. Simple memory: a theory for archicortex. *Proc. R. Soc. London Ser. B* 841:262;23–81

McClelland J, Goddard N. 1996. Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. *Hippocampus.* In press

McClelland J, McNaughton B, O'Reilly R. 1994. *Why we have complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory.* No. Tech. Rep. PDP.CNS.94.1. Carnegie Mellon Univ., Pittsburgh

McCulloch WS, Pitts W. 1943. A logical calculus of the ideas immanent in neural nets. *Bull. Math. Biographys.* 5:115–37

McNaughton B. 1989. Neuronal mechanisms for spatial computation and information storage. *Neural Connections, Mental Computations,* ed. L Nadel, L Cover, P Culicover, R Harnish, pp. 285–350. Cambridge, MA: MIT Press

McNaughton BL. 1991. Associative pattern completion in hippocampal circuits: new evidence and new questions. *Brain Res. Rev.* 16:202–4

McNaughton BL, Morris RGM. 1987. Hippo-

campal synaptic enhancement and information storage. *Trends Neurosci.* 10(10): 408–15
McNaughton BL, Nadel L. 1990. Hebb-Marr networks and the neurobiological representation of action in space. *Neuroscience and Connectionist Theory*, ed. M Gluck, D Rumelhart, pp. 1–63. Hillsdale, NJ: Erlbaum
Mishkin M. 1982. A memory system in the monkey. *Philos. Trans. R. Soc. London* 298:85–92
Moore J, Stickney K. 1980. Formation of attentional-associative networks in real time: role of the hippocampus and implications for conditioning. *Physiol. Psychol.* 8(2): 207–17
Moyer JR, Deyo RA, Disterhoft JF. 1990. Hippocampectomy disrupts trace eye-blink conditioning in rabbits. *Behav. Neurosci.* 104(2):243–52
Muller RU, Kubie JL, Ranck JB. 1987. Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *J. Neurosci.* 7(7):1935–50
Muller R, Stead M. 1996. Hippocampal place cells connected by hebbian synapses can solve spatial problems. *Hippocampus.* In press
Murre J. 1996. Models of amnesia. *Hippocampus.* In press
Myers C, Gluck M. 1994. Context, conditioning and hippocampal re-representation. *Behav. Neurosci.* 108(5):835–47
Myers CE, Gluck MA. 1996. Cortico-hippocampal representations in simultaneous odor discrimination: a computational interpretation of Eichenbaum, Mathews, and Cohen (1989). *Behav. Neurosci.* 110(4): 1–22
Myers CE, Gluck MA, Granger R. 1995. Dissociation of hippocampal and entorhinal function in associative learning: a computational approach. *Psychobiology* 23(2): 116–38
Myers CM, Ermita BR, Harris K, Hasselmo M, Solomon P, Gluck MA. 1996. A computational model of cholinergic disruption of septo-hippocampal activity in classical eyeblink conditioning. *Neurobiol. Learn. Mem.* 66:51–66
O'Keefe J, Nadel L. 1978. *The Hippocampus as a Cognitive Map.* Oxford: Clarendon
O'Reilly R, McClelland J. 1994. Hippocampal conjunctive encoding, storage, and recall: avoiding a tradeoff. *Hippocampus* 4:661–82
Otto T, Schottler F, Staubli U, Eichenbaum H, Lynch G. 1991. Hippocampus and olfactory discrimination learning: effects of entorhinal cortex lesions on olfactory learning and memory in a successive-cue, go-no-go task. *Behav. Neurosci.* 105(1):111–19
Penick S, Solomon R. 1991. Hippocampus context and conditioning. *Behav. Neurosci.* 105(5):611–17
Port R, Romano A, Patterson M. 1986. Stimulus duration discrimination in the rabbit: effects of hippocampectomy on discrimination and reversal learning. *Physiol. Psychol.* 4(3–4):124–29
Prepscius C, Levy W. 1994. Sequence prediction and cognitive mapping by a biologically plausible neural network. In *Proc. World Congr. Neur. Netw.* 4:164, 5:169. San Diego: INNS Press
Price J. 1973. An autoradiographic study of complementary laminar patterns of termination of afferent fiber to the olfactory cortex. *J. Compar. Neurol.* 150:87–108
Recce M, Harris K. 1996. Memory for places: a navigation model in support of Marr's theory of hippocampal function. *Hippocampus.* In press
Reilly S, Harley C, Revusky S. 1993. Ibotenate lesions of the hippocampus enhance latent inhibition in conditioned taste aversion and increase resistance to extinction in conditioned taste preference. *Behav. Neurosci.* 107(6):996–1004
Ribot T. 1882. *The Diseases of Memory.* New York: Appleton
Rolls E. 1989. The representation and storage of information in neural networks in the primate cerebral cortex and hippocampus. In *The Computing Neuron*, ed. R Durbin, C Miall, G Mitchison, pp. 125–59. Wokingham, UK: Addison-Wesley
Rosenblatt F. 1962. *Principles of Neurodynamics.* New York: Spartan
Rudy JW, Sutherland RJ. 1989. The hippocampal formation is necessary for rats to learn and remember configural discriminations. *Behav. Brain Res.* 34:97–109
Rudy JW, Sutherland RJ. 1995. Configural association theory and the hippocampal formation: an appraisal and reconfiguration. *Hippocampus* 5:375–89
Rumelhart D, Zipser D. 1985. Feature discovery by competitive learning. *Cogn. Sci.* 9: 75–112
Schmajuk NA. 1994. Stimulus configuration, classical conditioning, and spatial learning: role of the hippocampus. In *Proc. World Congr. Neur. Netw.* 2:723-28. San Diego: INNS
Schmajuk NA, Blair HT. 1993. Stimulus configuration, spatial learning and hippocampal function. *Behav. Brain Res.* 59: 103–17
Schmajuk NA, DiCarlo JJ. 1990. *Backpropagation, classical conditioning and hippocampal function.* Tech. Rep., Northwestern Univ.
Schmajuk NA, DiCarlo JJ. 1991. A neural network approach to hippocampal function in

classical conditioning. *Behav. Neurosci.* 105:82–110
Schmajuk NA, DiCarlo JJ. 1992. Stimulus configuration, classical conditioning and hippocampal function. *Psychol. Rev.* 99: 268–305
Schmaltz LW, Theios J. 1972. Acquisition and extinction of a classically conditioned response in hippocampectomized rabbits (*Oryctolagus cuniculus*). *J. Comp. Physiol. Psychol.* 79:328–33
Sharp P. 1991. Computer simulation of hippocampal place cells. *Psychobiology* 19(2): 103–15
Sharp P, Blair H, Brown M. 1996. Neural network modeling of the hippocampal formation spatial signals, and their possible role in navigation: a modular approach. *Hippocampus.* In press
Solomon PR, Moore J. 1975. Latent inhibition and stimulus generalization of the classically conditioned nictitating membrane response in rabbits (*Oryctolagus cuniculus*) following dorsal hippocampal ablation. *J. Comp. Physiol. Psychol.* 89:1192–203
Solomon PR, Solomon SD, Van der Schaaf E, Perry HE. 1983. Altered activity in the hippocampus is more detrimental to classical conditioning than removing the structure. *Science* 220:329–31
Squire L. 1987. *Memory and Brain.* New York: Oxford Univ. Press
Squire L, Butters N, eds. 1984. *Neuropsychol. Mem.* New York: Guilford
Squire LR, Alvarez P. 1995. Retrograde amnesia and memory consolidation: a neurobiological perspective. *Curr. Opin. Neurobiol.* 5:169–77
Squire LR, Cohen N. 1979. Memory and amnesia: resistance to disruption develops for years after learning. *Behav. Neur. Biol.* 25: 118
Sutherland R, Rudy J. 1989. Configural association theory: the role of the hippocampal formation in learning, memory and amnesia. *Psychobiology* 17(2):129–44
Teyler T, DiScenna P. 1986. The hippocampal memory indexing theory. *Behav. Neurosci.* 100(2):147–54
Treves A, Rolls E. 1992. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* 2(2):189–200
van Hoesen G, Pandya D. 1975. Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. I. Temporal lobe afferents. *Brain Res.* 95:1–24
Volpe BT, Hirst W. 1983. Amnesia following the rupture and repair of an anterior communicating artery aneurysm. *J. Neurol. Neurosurg. Psychiatry.* 46:704–9
Weiskrantz L, Warrington EK. 1979. Conditioning in amnesic patients. *Neuropsychologia* 17:187–94
Wickelgren W. 1979. Chunking and consolidation: a theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychol. Rev.* 86(1):44–60
Willshaw D, Buckingham J. 1990. An assessment of Marr's theory of the hippocampus as a temporary memory store. *Philos. Trans. R. Soc. London Ser. B* 329:205–15
Wilson MA, McNaughton BA. 1994. Reactivation of hippocampal ensemble memories during sleep. *Science* 265:676–79
Woodhams PLL, Celio MR, Ulfig N, Witter MP. 1993. Morphological and functional correlates of borders in the entorhinal cortex and hippocampus. *Hippocampus* 3: 313–11
Woodruff-Pak D. 1993. Eyeblink classical conditioning in HM: delay and trace paradigms. *Behav. Neurosci.* 107(6):911–25
Zola-Morgan, Squire LR. 1993. Neuroanatomy of memory. *Annu. Rev. Neurosci.* 16:547–63