HIPPOCAMPUS 6:643–653 (1996)

Integrating Behavioral and Physiological Models of Hippocampal Function

Mark A. Gluck and Catherine E. Myers

Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, New Jersey

ABSTRACT: In recent modeling of hippocampal function, we have attempted to integrate formal behavioral analyses of classical conditioning with psychobiological data on brain lesions (Gluck and Myers [1993] Hippocampus 3:491–516; Myers and Gluck [1994] Behav Neurosci 108(5):835–847). Based on comparative behavioral analyses, we have argued that animals with hippocampal region damage are unable to alter stimulus similarity based on experience. While hippocampal-damaged animals can still learn whether to respond to an individual stimulus, they are notably impaired at many tasks involving learning relationships between stimuli—especially in the absence of explicit reinforcement. These analyses lead to a computational theory which identifies two representational recoding processes—*predictive differentiation* and *redundancy compression*—which alter stimulus similarity relationships in intact animals but are dependent on intact hippocampal region processing. More recent, and ongoing, modeling aims to broaden this model of hippocampal region function in classical conditioning, with an emphasis on physiological and anatomical constraints, including the role of the fornix and subcortical modulation, preprocessing in sensory cortices, and localization of the proposed representational functions within more precisely identified hippocampal region substrates (Myers et al. [1995] Psychology 23(2):116–138; Myers and Gluck [1996] Behav Neurosci; Myers et al. [1996] Neurobiol Learning Memory). Working to bridge between behavioral and physiological levels of analysis, we ultimately hope to develop a more complete understanding of hippocampal region function in memory across a wider range of behavioral paradigms, elucidating how this functionality emerges from underlying physiological and anatomical substrates. © 1997 Wiley-Liss, Inc.

KEY WORDS: entorhinal cortex, cholinergic, computational, amnesia

INTRODUCTION

What are the Data to be Explained?

It is widely accepted that the hippocampal region, including hippocampus, dentate gyrus, subiculum, and entorhinal cortex (see Fig. 1) plays some critical role in learning and memory, but a precise characterization of that role remains elusive. Over the years, a large body of data has evolved detailing which kinds of tasks are impaired and which are spared following hippocampal region damage. The theoretical challenge has been to map these disparate and varied data to some underlying functional interpretation of the hippocampus. While there has been considerable theoretical and empirical progress, the very breadth of this data has led to many confusing and often contradictory conclusions.

One impediment to theoretical convergence in theories of hippocampal function has been the large number of species, preparations, and behavioral paradigms that have been studied. In an effort to simplify matters, some researchers have begun by focusing on classical (Pavlovian) conditioning, an elementary associative learning paradigm for which the neural bases are relatively well understood (Thompson, 1986, 1988). Briefly, classical conditioning involves the repeated pairing of a previously neutral cue (the conditioned stimulus or CS) with a response-evoking cue (the unconditioned stimulus or US) until the CS alone evokes an anticipatory response (the conditioned response or CR). For example, the US may be a corneal airpuff which evokes a protective blink; if the airpuff is repeatedly paired with a tone or light, that CS eventually evokes anticipatory blinks (Gormezano et al., 1983). One advantage of this domain is that it is well studied in intact animals, and several detailed mathematical and computational models exist which account for a wide range of conditioning behaviors (Sutherland and Mackintosh, 1971; Rescorla and Wagner, 1972; Pearce and Hall, 1980; Mackintosh, 1983). A second advantage to classical conditioning as a model system for the studying the neural bases of memory is that the biological substrate is well understood, especially for motor-reflex conditioning where the necessary and sufficient cerebellar circuits have been delineated (e.g., Thompson, 1986, 1990). The third advantage to studying classical conditioning is that although the hippocampus is not necessary for this simplest form of associative learning in animals (Schmaltz and Theios, 1972; Solomon and Moore, 1975; Port and Patterson, 1984; etc.) or humans (Daum et al., 1989; Woodruff-Pak, 1993; Gabrieli et al., 1995), more complex conditioning paradigms involving con-

Accepted for publication July 15, 1996.
Address correspondence reprint requests to Mark A. Gluck, Ctr. for Molecular and Behavioral Neuroscience, Rutgers Univ., 197 University Avenue, Newark, NJ 07102.

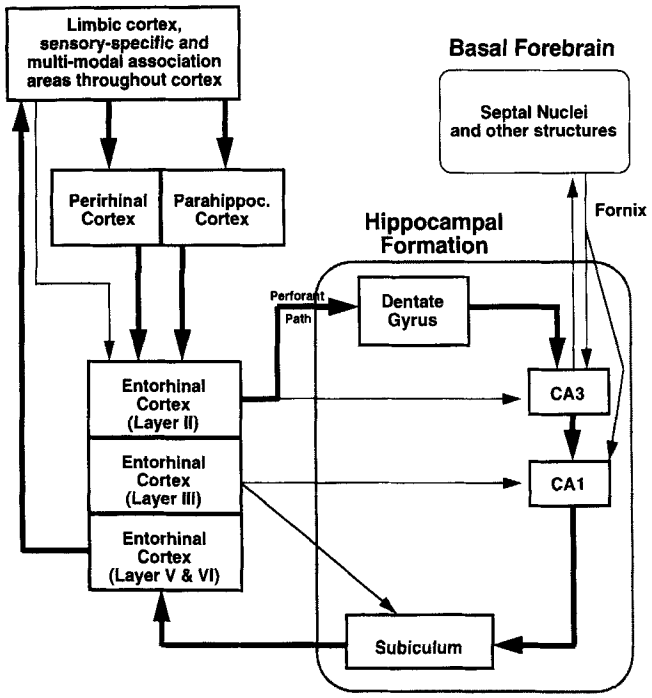

FIGURE 1. Schematic of major information flow pathways in the hippocampal region. Figure reprinted from Myers and Gluck (1996).

figural, contextual, or temporal associations are known to be disrupted. Furthermore, even during the simplest forms of conditioning, the hippocampus is active, as evidenced by neurophysiological recordings which show changes in hippocampal activity to reflect learning (Berger and Orr, 1983; Sears and Steinmetz, 1990).

Of particular interest to theory and model development is the fact that the extent of impairment from damage to the hippocampal region often depends critically on precise lesion extent (Honey and Good, 1993; Reilly et al., 1993; Otto et al., 1991; Jarrard, 1993; etc.). These and other results suggest 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.

For these reasons we have chosen classical conditioning as a well-defined domain within which to begin to model the functional role of the hippocampal region. The data we seek to explain are the patterns of classical conditioning performance on a variety of conditioning paradigms in both intact subjects and after damage to the hippocampal region as a whole or damage limited to one or more of the substructures.

What Is the Goal of the Model?

One goal of the model, as stated above, is to capture the behavior of intact and lesioned subjects on a wide variety of conditioning paradigms. A model which can do this will help shed light on what function the hippocampal region could be computing, and will generate predictions about what new behaviors ought to be impaired and spared after hippocampal region damage.

A second goal of the model is to generate these behaviors using mechanisms which are consistent with what we know about hippocampal region anatomy and physiology. There may be many ways in which to produce the desired behaviors; the only way to distinguish among these models is to determine which postulated processes are plausible as emergent capabilities of the anatomical substrate.

How to Evaluate the Model?

This kind of model should be evaluated as to how well it fulfills its two goals noted above and as to whether it does so better than other existing models. The model should capture as broad a range of behavioral data as possible, and this range should be compared both to the range of data accounted for by other behavioral models of (intact) conditioning (e.g., Rescorla and Wagner, 1972; Pearce and Hall, 1980; Mackintosh, 1983) and to other models which address hippocampal lesion effects on conditioning (e.g., Buhusi and Schmajuk, 1996, this issue). Finally, the model should only postulate processes which are consistent with known anatomical and physiological substrates.

The true test of the model is its ability to make testable novel predictions. In addition to providing a way to falsify the model, these predictions can also be used to guide empirical research toward potentially interesting new directions.

A final level of evaluation is qualitative: Does the model provide a useful way of thinking about hippocampal function? Does the model give us a clearer way of picturing hippocampal region function than we had before? Does the model provide a simple and intuitively compelling interpretation for a wide range of behavioral and biological data?

STIMULUS REPRESENTATION AND THE HIPPOCAMPUS

In Gluck and Myers (1993), we presented a computational theory of hippocampal region function which focused on stimulus representations and how they evolve—or adapt—during associative learning. A stimulus representation might be the pattern of activities evoked by that stimulus over a group of neurons in the brain or nodes in a connectionist network. Learning to map various stimuli to responses can be facilitated by appropriate stimulus representations. For example, if two stimuli evoke very similar representations, there will be a high tendency to generalize (i.e., transfer) what is learned about one to the other. If those two stimuli are to be mapped to different responses, however, it would be more helpful if their representations were easily differentiated. In Gluck and Myers (1993) we proposed that the hippocampal region is able to alter the degree of generalization between stimuli by altering their stimulus representations according to two specific principles. The first, termed "predictive differentiation," is a bias to differentiate (or reduce similarity between) representations

of stimuli which predict different future events (such as responses). The second, termed "redundancy compression," is a complementary bias to compress (or increase similarity between) representations of stimuli which predict similar future events, or which reliably concur.

This function can be implemented in a connectionist network such as that shown on the right in Figure 2A (Gluck and Myers, 1993). That network is an autoencoder (Hinton, 1989), which learns to map from stimulus inputs to outputs which reconstruct those inputs as well as predicting the correct behavioral response. An autoencoder differs from the autoassociative networks which have often been proposed to model hippocampus (Marr, 1971; McNaughton, 1989; Treves and Rolls, 1992; Hasselmo et al., 1995; etc.) in that it includes a narrow internal or hidden layer of nodes. Because this layer cannot contain all the information present in the inputs, the network must construct representations of the inputs in the internal layer which compress redundancies while preserving and differentiating predictive information. Thus, these representations are biased by the same constraints proposed to operate in the hippocampal region. It is important to distinguish at the outset between this model (and the backpropagation algorithm of Rumelhart et al. [1986], used to train it) and the qualitative theory: There may be many ways to implement our general theory which accomplish the same basic function. In later sections we will return to this idea and show how there may be more physiologically realistic ways to implement the putative hippocampal-dependent stimulus recodings.

As they are evolving, stimulus representations constructed in the hippocampal region model are assumed to be made available to other cortical and cerebellar regions which are the sites of long-term memory storage, but which cannot themselves alter representations in the same way that the hippocampal region can. Figure 2A shows a complete intact model, including one such cortical network. The hippocampal region acts as a "teacher" for the cortical network, providing the desired activations of the hidden layer nodes in the cortical network, which can be acquired by application of a simple learning rule such as the error-correcting least mean squared (LMS) learning rule (Widrow and Hoff, 1960), which is closely related both to psychological learning rules (Rescorla and Wagner, 1972) and biological plasticity mechanisms such as long-term potentiation (LTP; Levy et al., 1983). A second application of LMS is then used to train the cortical network to map from these acquired representations to output activations which are the system's behavioral response. All learning in the model takes place incrementally. Thus, while the model does not specifically address consolidation of hippocampal memories to long-term store (as do, e.g., the theories of McClelland and Goddard, 1996; Murre, 1996 in this issue), our model is broadly consistent and is certainly compatible with the idea that there may be an indeterminately long period before hippocampal information is completely acquired by the neocortex.

Within Gluck and Myers's cortico-hippocampal model hippocampal region damage is simulated by disabling the hippocampal region network on the right in Figure 2A, resulting in the reduced network shown in Figure 2B. In this simulation of hippocampal lesion, the cortical network is assumed to be no longer able to modify its representations in the hidden layer. Within the model, this is instantiated by assuming that the lower layer of cortical weights are be fixed. The cortical network can, however, still learn new behavioral responses by mapping from its pre-existing (and now fixed) representations to appropriate outputs.

The intact and lesioned models of Figure 2 can be applied to simple associative learning paradigms such as classical conditioning, by assuming that the inputs represent the presence or absence of CSs and contextual cues, while the output is trained to anticipate US arrival. The model captures many trial-level behaviors in both intact and lesioned animals and also generates novel predictions about what kinds of behaviors ought to be im-

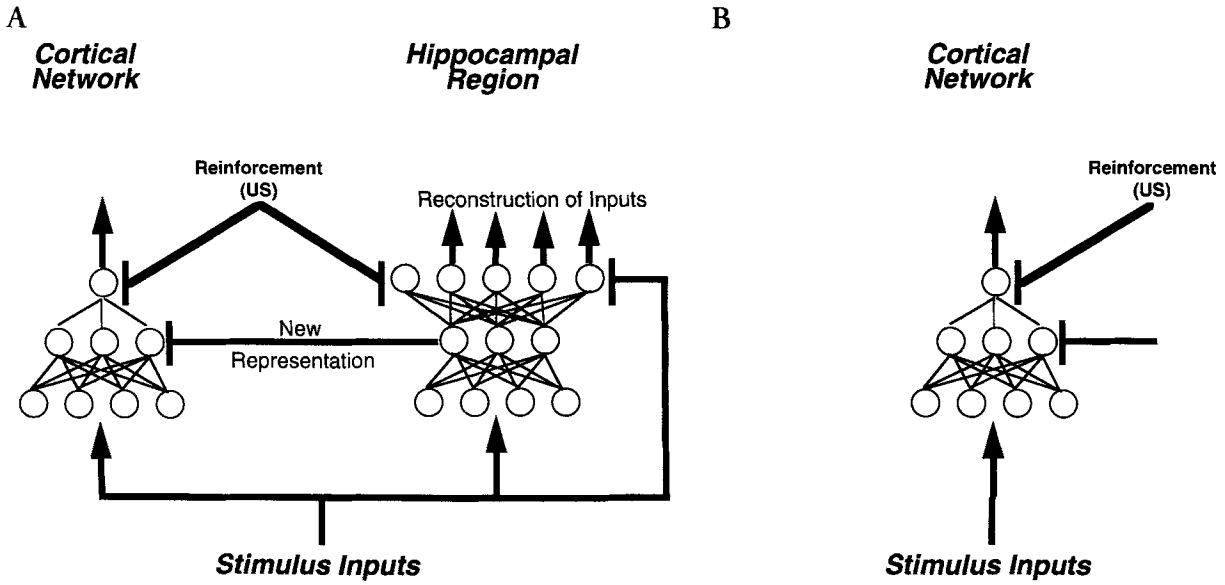

FIGURE 2. The cortico-hippocampal model (Gluck and Myers, 1994). A: Intact model. B: Hippocampal-lesioned model. Figures reprinted from Myers and Gluck (1996).

paired or spared after hippocampal region damage (Gluck and Myers, 1993; Myers and Gluck, 1994, 1996). We review a few of these in the next section.

First, though, we note that our theory is based on the assumption that hippocampal-dependent representational processes result in specific changes that should be reflected in neuronal activity patterns. There is some neurophysiological evidence that such effects do occur during learning. For example, our theory expects that the representation of co-occurring (redundant) stimuli should gradually become compressed, or more similar. Sakai and Miyashita (1991) recorded neuronal responses in anterior temporal cortex of monkeys trained on a paired associate task and found that paired stimuli do indeed come to elicit significantly correlated responses. Conversely, our theory expects that the representations of stimuli which are mapped to different responses should gradually become differentiated, or less similar. Cahusac et al. (1993) recorded from hippocampal and parahippocampal cells during discrimination tasks in the monkey and found that neurons which initially responded equally to the stimuli altered their responses to differentiate stimuli as the task was learned. Thus, there is neurophysiological evidence to support the proposed function; it remains to be seen whether these neuronal changes are dependent on the hippocampal region, as the theory predicts.

Stimulus Representations and Classical Conditioning

Consistent with empirical data, the cortico-hippocampal model of Figure 2 predicts that there should be no specific impairment for simple conditioning after hippocampal lesion. Figure 3A shows that the lesioned model learns a simple discrimination (e.g., respond to stimulus A but not stimulus B) as fast as the intact model (Gluck and Myers, 1993). Although the lesioned model can no longer adaptively modify stimulus representations in the same way that the intact model can, a simple discrimination of highly differentiable stimuli (such as a tone and a light) is not expected to require any new representation in order that the animal may respond to one, but not the other: any representation which at least partially distinguishes A and B is probably sufficient, and all the lesioned model has to do is learn to map from these prior fixed representations to different responses.

However, if the task becomes more complex, the ability to modify representations becomes critical. One paradigm which has received much attention is latent inhibition (Lubow, 1973), which refers to the fact that normal animals and humans are slower to learn to respond to a stimulus A if they have received prior unreinforced exposure to A. Conceptually, it appears that the preexposure teaches subjects that A is "unimportant," and may safely be "tuned out" of attention. In the later phase of training, when A does become important, the attention to A must be explicitly "tuned back in" before the response can be learned, and this results in slower learning compared to subjects who did not receive exposure. The intact model correctly shows this effect (Fig. 3B; Gluck and Myers, 1993). During the exposure phase, A is presented repeatedly in the context of whatever background cues are present—such as the sights and sounds of the experimental chamber. Because these contextual cues co-occur with A, and because they, like A, predict no particular salient future events, the hippocampal region network tends to compress the contextual cues with the experimental cue, A. This compression of the cue and context leads to difficulties in the subsequent phase, when the task is to learn to respond to A (in the context) but not to the context alone. As a result, learning after pre-exposure to A is slower relative to a control condition exposed to the context alone (Fig. 3B; Myers and Gluck, 1994). Because the latent inhibition effect in our model depends on hippocampal network–mediated compression, it is absent in the lesioned model. Consistent with this model expectation, broad hippocampal region damage eliminates latent inhibition in animals (Kaye and Pearce, 1987; Solomon and Moore, 1975) although, as discussed in the next section, more selective lesions do not.

In the same way, redundancy compression mechanisms can be used to generate other behavioral effects in the intact model, such as sensory preconditioning and compound preconditioning; predictive differentiation effects are used to generate such effects as facilitated reversals after overtraining or successive reversals, and easy-hard transfer (Gluck and Myers, 1993). The fact that these effects are eliminated in the lesioned model leads to predictions that they should likewise be hippocampal-dependent in animals. Many of these predictions are consistent with existing data, but some lead to novel predictions which remain to be tested empirically.

Modeling Contextual Sensitivity

The Gluck and Myers model of Figure 2 also applies to data regarding the effects of hippocampal lesion on contextual processing. Many of the learning deficits associated with hippocampal damage can be described as contextual deficits, as they suggest an inability to incorporate information about the environmental conditions under which an event occurs (see Hirsh, 1974; Nadel and Wilner, 1980). In fact, the latent inhibition effect described above, and its disruption after hippocampal region damage, are interpreted as contextual effects in the model (Myers and Gluck, 1994). It should be noted that hippocampal damage does not result in a general inability to perceive contextual cues, since lesioned animals can still learn to discriminate contexts (e.g., Good and Honey, 1991; Phillips and LeDoux, 1994); the lesioned model correctly maintains this ability as well (Myers and Gluck, 1994). Instead, what seems to be disrupted in the lesioned animal is the ability to use context to help interpret the meaning of cues in specific situations (Myers and Gluck, 1994). Within the intact model, contextual information is included in the representations of the cues—since the context always co-occurs with the cue. Therefore, learning about a cue is always at least partially context dependent. As a result, if the cue is presented in a novel context, there is a decrease in strength of responding (Fig. 3C). Because this effect is dependent upon hippocampal network–mediated representational changes, it is not seen in the lesioned model. Similarly, under some circumstances, a response decrement is seen after context shift in intact animals, and the effect is eliminated after hippocampal region damage (Honey and Good, 1993). The model can account for a range of contextual effects

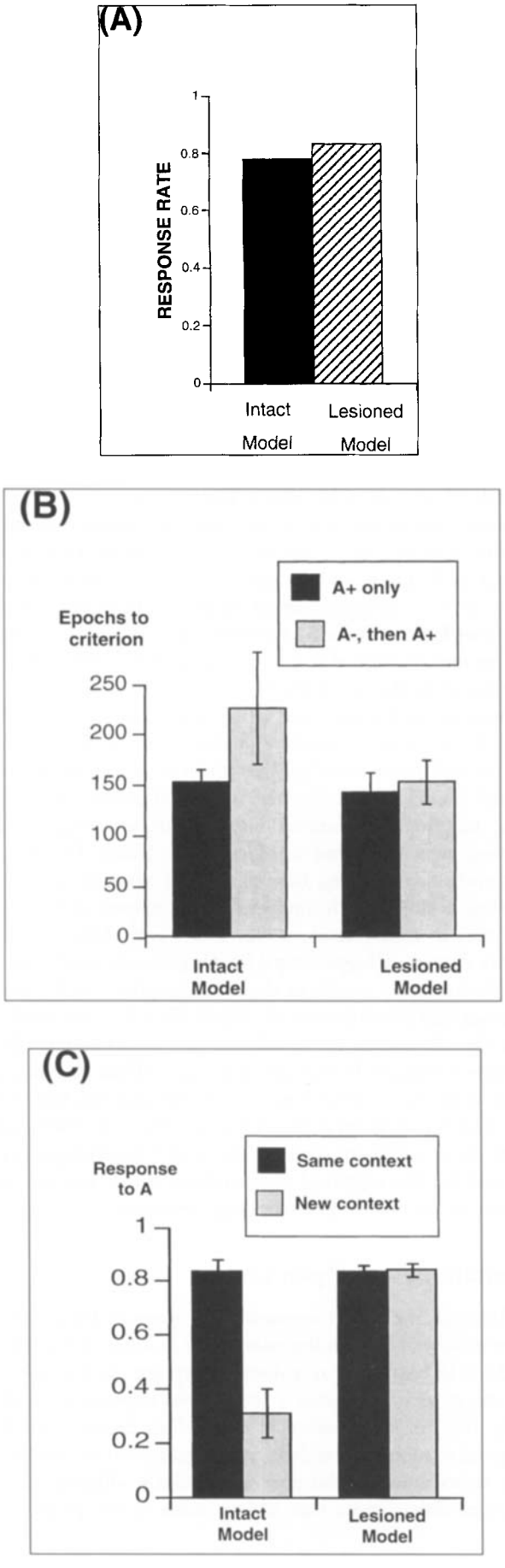

FIGURE 3. Simulations results with the intact and lesioned models. A: Simple discrimination of two cues, A and B, is not slowed by hippocampal region damage in the model. Figure reprinted from Myers and Gluck (1996). B: Latent inhibition, retarded learning to respond to A after prior exposure to A in the same context, is correctly shown in the intact but not the lesioned models. Figure reprinted from Myers et al. (1995). C: In the intact model a response trained to a cue in one context drops when that cue is presented in a new context; this response decrement after context shift is not seen in the lesioned model. Figure reprinted from Myers et al. (1995).

in this way and also provides a computational instantiation of several existing qualitative theories which implicate the hippocampus in contextual learning (e.g., Hirsh, 1974; Penick and Solomon, 1991; Nadel and Wilner, 1980).

Extensions to Other Forms of Learning

The data discussed above are from classical conditioning studies, especially the eyeblink response preparation. But the basic explanatory mechanism of the model is its use of modifiable representations and of biases to compress and differentiate those representations based on environmental contingencies. These processes are expected to be found in many different kinds of learning, not just eyeblink conditioning. For example, we have applied the model to a simple operant paradigm, the simultaneous odor discrimination studied in rats by Eichenbaum and colleagues (Eichenbaum et al., 1988, 1989). In this task, the rat is presented with two odors, and must learn to approach the positive odor to obtain a reward. Normal rats learn such discriminations within a few hundred trials and show savings effects facilitating subsequent similar problems; fornix-lesioned rats are greatly impaired (Eichenbaum et al., 1988). To implement such a task in the model requires some elaborations, such as a simple model of piriform cortex, adapted from the physiologically realistic model of Ambros-Ingerson et al. (1990), and the ability to generate multiple outputs, instead of just a single conditioned response (see Myers and Gluck, 1996, for further details). However, these changes do not alter the basic processes attributed to the hippocampal region, or the relationship assumed to exist between hippocampal region and cortical storage sites. The resulting intact and lesioned models account for many aspects of the data presented by Eichenbaum et al., 1988, 1989 including the relative difficulty of simultaneous discrimination in lesioned rats, the occasionally good performance of individual lesioned rats on particular problems, and the ability of intact but not lesioned rats to transfer to new discriminations using previously learned odors (Myers and Gluck, 1996). Notably, the reasons for these behaviors in the model are quite similar to those posited by Eichenbaum et al., 1988, 1989 to explain their empirical data (see also Gluck and Myers, 1995).

Connections to Biological Substrate

One of the goals of this modeling work was to address as broad a range as possible of associative learning behaviors in intact and lesioned models, and the work described above attempts to do this. A second important issue is whether the postulated functions

can be implemented in behaviorally plausible ways. There is good evidence that the cerebellum, the critical substrate for classical eyeblink conditioning (Thompson, 1986, 1988), does perform the kind of simple error correction assumed in the cortical network of our model (see Gluck et al., 1994a,b). However, the hippocampal region network has an architecture and learning algorithms which are not obviously related to the substrate (although Schmajuk and DiCarlo, 1990, suggest that a version of backpropagation is biologically plausible).

We would prefer to have a hippocampal region model which incorporates anatomical constraints more directly than in the original Gluck and Myers (1993) specification. Most hippocampal models which incorporate anatomical constraints have assumed that the hippocampus is self-organizing—learning without reliance on outside error information (see Grossberg, 1976; Kohonen, 1984; Rumelhart and Zipser, 1985). To date, we have shown that this kind of self-organization is sufficient to account for at least part of the representational processing the model attributes to the hippocampal region as a whole. In particular, we started with the assumption made by Ambros-Ingerson and colleagues (1990) that the anatomy and physiology of the superficial piriform cortex is sufficient to implement a competitive network model, which can learn to form hierarchical clusters of odor inputs. Because the piriform cortex and entorhinal cortex elide in rat, and their superficial layers are closely related anatomically and physiologically, it is possible that the entorhinal cortex performs a similar function (Myers et al., 1995; Gluck and Granger, 1993)—although the entorhinal cortex presumably operates on a broad spectrum of polymodal and crossmodal stimulus features, while the piriform cortex is primarily concerned with odors. Most importantly, such a network performs compression of concurring stimuli, one of the functions we previously proposed occurs in the hippocampal region.

Therefore, we constructed a reduced model, with an "entorhinal" network which performs stimulus-stimulus compression to replace the full hippocampal region network (Fig. 4A; Myers et al., 1995). This model is compared to a lesion which selectively damages the hippocampus and dentate gyrus, but leaves intact the entorhinal cortex. In animals, 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 and Good, 1993; Reilly et al., 1993). The selectively lesioned model produces the same effect (Fig. 4B: 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 making specific 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. Empirical studies are now being conducted to test these predictions.

Connections to Hasselmo's Cholinergic Model

In addition to the information-carrying pathways between hippocampus and neocortex, there is also an important bi-directional pathway through the fornix connecting the hippocampus with subcortical structures (refer to Fig. 1). One important input through the fornix is a modulatory cholinergic input from the medial septum. Hasselmo and Schnell (1994) have suggested that this cholinergic input can be used to switch the hippocampus between two processing states: information storage, assumed to take place in the presence of acetylcholine (ACh), and information retrieval, assumed to take place in the absence of acetylcholine. Hasselmo and Schnell argue that such a separation of storage and retrieval dynamics is necessary to prevent runaway synaptic modification during storage (see article in this issue). The hypothesis is also consistent with behavioral data showing that classical conditioning is retarded after interruption of septohippocampal cholinergic inputs by medial septal lesion (Berry and Thompson, 1979) or the anticholinergic drug scopolamine (Fig. 5A; Solomon et al., 1983; Solomon et al., 1993).

We have implemented a simplified version of Hasselmo's cholinergic hypothesis within our intact cortico-hippocampal model. This is done by noting 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 our model by lowering this learning rate—although *not* the rate at which this information is transferred to the cortical network or the rate at which cortical associations develop. The consequence of this depressed hippocampal learning rate is to prolong the initial nonresponding phase prior to onset of the initial conditioned responses (Fig. 5B), much as is seen in the experimental data (Fig. 5A). 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 (see Myers et al., 1996, for review). 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 (see Myers et al., 1996, for review). The model therefore provides an account for this empirical phenomenon, which has been problematic in the clinical pharmacology literature.

Limitations and Open Issues

Although the model accounts for a range of behavioral data, there are several obvious limitations. For example, it is a trial-level model, which means that it does not capture any within-trial information, such as whether a conditioned response is timed correctly, nor can it be applied to tasks which depend critically on temporal relationships, such as trace conditioning or conditioning with altered interstimulus intervals. Similarly, although the work described above shows that at least some of the proposed hip-

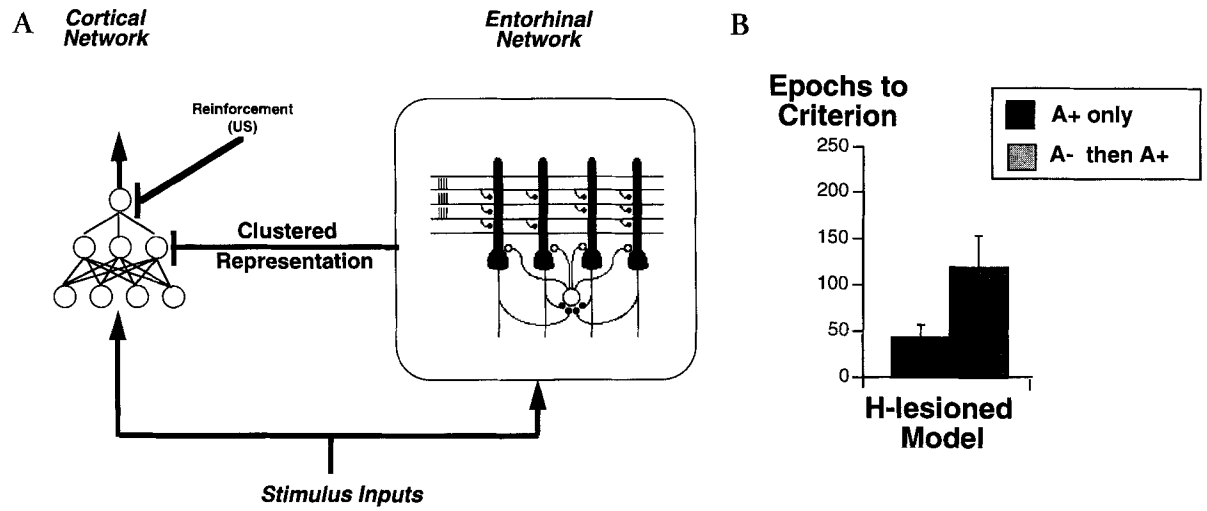

FIGURE 4. A: The hippocampal region network can be replaced by one representing biologically plausible stimulus compression in the entorhinal cortex; this approximates the effects of selective hippocampal lesion which spares entorhinal processing. B: Such a model correctly maintains latent inhibition, even though a broader lesion (and the fully lesioned model) do not (refer Fig. 3B). Figure adapted from Myers et al. (1995).

pocampal region processes are consistent with the anatomy and physiology, it remains to be shown that the full functionality can be so accounted for, and that these subfunctions interact in reasonable ways. Finally, although the model makes specific, testable predictions, many of these predictions remain to be tested, and such tests which will ultimately determine the validity of the model.

DISCUSSION

What Do We Know Now That We Did Not Know Before?

One of the most important goals of our modeling work has been to show that a simple description of a putative information-processing role of the hippocampal region in associative learning is sufficient to account for a range of associative learning data from intact and lesioned animals. To the extent that this endeavor succeeds, it provides a simple and intuitive description of what the hippocampus is computing. In many cases, this basic theory is sufficient to allow for the understanding of why phenomena should be hippocampal dependent or independent. For example, the discussion of latent inhibition above relies more on the qualitative explanation of the theory than on the details of specific model simulations. Other phenomena are amenable to the same treatment.

In addition to providing a general framework for understanding hippocampal region function, our model provides an understanding of how several previously problematic phenomena might occur: These include the elimination of latent inhibition after a large hippocampal region lesion but not a smaller hippocampal-only lesion, the impaired learning after hippocampal disruption but not outright hippocampal removal, the improved learning after a moderate dose of cholinergic agonists but not after a larger dose, and so on. The fact that these explanations follow easily from the model, but were not apparent before, illustrates one way in which this kind of modeling work can be useful.

The model makes several novel predictions which can be used to test the model's validity. These kinds of prediction are currently guiding new empirical research, by suggesting avenues of exploration which may prove particularly useful in elucidating hippocampal function—whether or not the specific predictions of the model are proven, or disproven, by the results.

What Did the Model Accomplish That Could Not Have Been Accomplished by Simpler Verbal-Qualitative Reasoning?

As argued above, many of the implications of our representational theory of hippocampal function can, in fact, be deduced at a qualitative level of reasoning. However, instantiating these representational processes as a computational model has two primary purposes. First, it allows quantitative evaluation of the theory. It allows consideration of parameter dependence, and it also illustrates subtle interactions within complex behavioral paradigms which are not always immediately obvious from deductive reasoning alone.

The second advantage of the computational model is that it provides an existence proof for the theory. It does not prove that the brain actually operates in the manner described. But, it does prove that the proposed mechanisms are sufficient to generate the desired behaviors. Further, as the top-down processing is implemented via bottom-up modules, the modeling provides an ever-stronger argument that the theory is, at least, a plausible description of how these brain regions might interact and process information. This strengthens the theory, compared with other theories that are not computationally implemented.

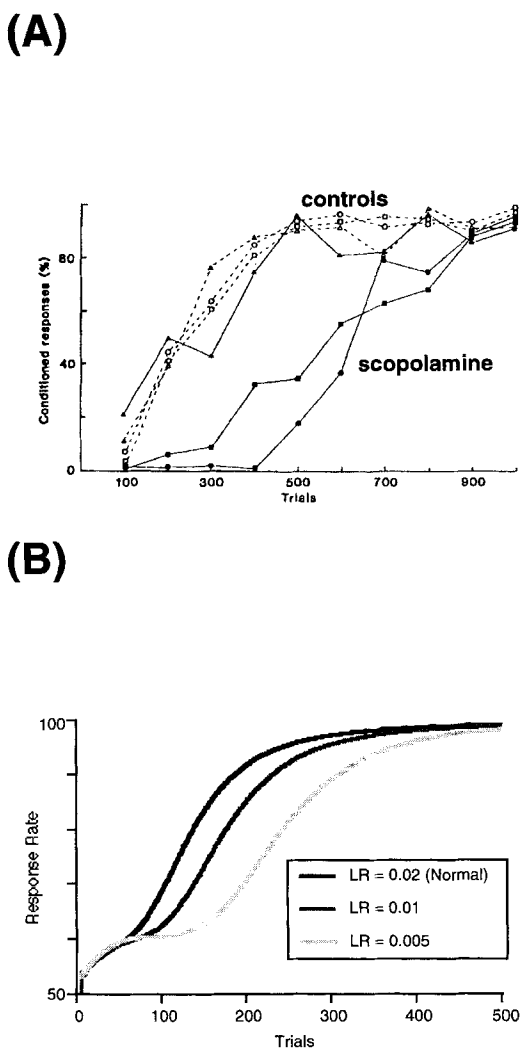

FIGURE 5. The anticholinergic drug scopolamine produces a dose-dependent retardation of eyeblink conditioning in animals and humans, specifically by prolonging the initial phase prior to onset of conditioned responding. B: A similar effect is obtained in the intact model by reducing the hippocampal learning rate, consistent with Hasselmo and colleagues' cholinergic hypothesis. Figures reprinted from Myers and Gluck (1996).

How Does This Model Relate to Others in This Issue and to Other Noncomputational Theories That Have Been Proposed?

In the earlier section, Stimulus-Representation and the Hippocampus, we have already noted many connections between our work, and other existing computational models and qualitative theories. In this section, we note a few other comparisons that have not been fully described above.

Psychobiological theories of hippocampal function in conditioning

The modeling presented here is most directly comparable in scope and aims to that of Schmajuk and colleagues. In Myers et al. (1995) we discussed the similarities and important differences between our model and earlier models by Schmajuk and DiCarlo (1990, 1992). One of the most important differences is that while our model assumes that extrahippocampal sites such as cerebellum are sufficient for simple error-correction learning, Schmajuk and DiCarlo's model assumes that the hippocampus is necessary for cue competition. Thus, while our model predicts that such effects as blocking, conditioned inhibition, and overshadowing should be hippocampal independent, the Schmajuk-DiCarlo model assumes they should be eliminated after hippocampal damage. The currently existing empirical data are contradictory and fail to adequately discriminate the models (see Myers et al., 1995, for review).

In their contribution to this issue, Buhusi and Schmajuk (1996) present a model of hippocampal function in conditioning that attributes attentional and configural mechanisms to specific components of the hippocampal region. Buhusi and Schmajuk propose 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 overlaying cortices are essential for *stimulus-stimulus redundancy compression*. This is consistent with studies showing that latent inhibition, a result we have interpreted as being mediated by stimulus compression, is spared after hippocampal lesions which do not extend to entorhinal cortex (Honey and Good, 1993; Reilly et al., 1993).

A second main difference between the two models is how they interpret the functional role of the medial septal inputs to the hippocampus. Buhusi and Schmajuk interpret the septohippocampal cholinergic pathways as providing an *error signal* that drives learning. In contrast, we have argued that these pathways can be functionally interpreted as providing *modulation of learning rates*, building upon similar arguments by Hasselmo (see Hasselmo et al., 1996, in this issue). Despite different functional interpretations of the medial septal inputs, both models correctly expect that cholinergic antagonists (such as scopolamine) should impair acquisition, but not latent inhibition. Buhusi and Schmajuk have not, however, addressed the detailed aspects of learning curves which are analyzed by Myers et al. (1996), nor the dose-dependency effects which lead to degraded learning in the presence of large amounts of acetylcholine.

Thus, although there are many superficial similarities between the two models, they differ greatly in specifics and in predictions. Clearly, further empirical tests are needed to determine which model accounts more fully for behaviors in intact and variously lesioned animals.

Theories and models of consolidation and human amnesia

The Gluck and Myers (1993) model was strictly limited in scope to address trial-level aspects of associative learning paradigms. This reflected an intentional initial limiting of scope, as we feel strongly that it is advisable to explore one domain in depth before trying to widen the applicability of the model. In this domain, the incremental nature of learning and the fact that most lesions are done prior to any training, we have not addressed the data regarding consolidation and temporally graded retrograde amnesia. Several connectionist models have considered this do-

main and generally suggest that the consolidation period represents incremental adoption by cortical storage sites of information initially captured in the hippocampus (e.g., Alvarez and Squire, 1994; Murre, 1994; McClelland et al., 1995; see also articles by Murre and McClelland and Goddard in this issue). Although we have not explicitly considered retrograde amnesia within our model, we have noted that incorporating a delay in transfer between hippocampal and cortical networks in our model might be the basis for incorporating consolidation into our model (Gluck and Myers, 1993).

Other qualitative theories of hippocampal function

Eichenbaum, Cohen, and colleagues' (Eichenbaum et al., 1992a,b) suggestion that the hippocampus is needed to form flexible representations during learning is related to our demonstration that the absence of appropriate stimulus representations during learning can result in altered transfer performance in hippocampal-lesioned animals (Gluck et al., 1994b). We have already noted above how our model instantiates qualitative suggestions that the hippocampal region is necessary for contextual processing—even though our model does not assume context per se is the hippocampus's chief domain (cf., Hirsh, 1974; Penick and Solomon, 1991).

Similarly, our model assumes that configural learning may often require hippocampal-dependent stimulus-stimulus learning and may therefore be especially susceptible to hippocampal damage (Myers and Gluck, 1994). This is consistent with Rudy and Sutherland's (1989) theory that configural learning is especially hippocampal dependent (see also Alvarado and Rudy, 1995, for discussion of this convergence).

For the same reason, our model stands in stark contrast to qualitative theories that the hippocampus is involved primarily in spatial learning (e.g., O'Keefe and Nadel, 1978). Clearly, spatial learning is extremely disrupted in animals with hippocampal region damage (e.g., Morris, 1983; Jarrard, 1993); it is also true that "place cells" form in the hippocampus which respond preferentially when the animal is in a particular region of space (e.g., O'Keefe, 1979; McNaughton et al., 1991). However, our model and theory suggest that the hippocampus is involved in all kinds of learning, and that those tasks which depend heavily on new representations are most likely to be disrupted by hippocampal region damage. For this reason, our theory predicts that spatial learning, which presumably involves associating arbitrary views and proprioceptive information into concepts of "place," might be especially sensitive to hippocampal damage—even though spatial learning per se is not assumed to be the hippocampus's function. This argument is similar to that advanced by many others who have considered possible information-processing roles for the hippocampus (e.g., Eichenbaum et al., 1988; Taube, 1991).

Finally, Bunsey and Eichenbaum (1993) have suggested that the entorhinal cortex (or parahippocampal region in general) mediates the "fusion" of concurring or nearly coincident stimuli; this process is functionally identical with the redundancy compression function we propose (see discussion in Gluck et al., 1994a,b).

What New Experimental Directions Are Suggested by This Modeling?

As noted earlier, our original model makes several novel predictions regarding behavioral effects which should be hippocampal dependent and independent. The suggestion that the entorhinal cortex can mediate stimulus compression leads to a similar round of predictions that those hippocampal-dependent effects which we assume depend on that mechanism should survive a lesion strictly limited to the hippocampus. To date, there are two results showing that latent inhibition does survive such a selective lesion although it is disrupted by broader hippocampal region damage. Finally, the analysis of septohippocampal cholinergic mediation predicts that cholinergic disruption should slow—but not eliminate—hippocampal processing. Thus, hippocampal-dependent behaviors should be retarded but still observed after such a disruption. This is consistent with the finding that scopolamine does not eliminate latent inhibition (Moore et al., 1986; see also Weiner, 1990). The remaining predictions all remain to be tested in animals. We are currently establishing an animal laboratory to examine some of these predictions empirically.

Because the model explains hippocampal region function in terms of information processing, instead of task-specific or species-specific roles, it does not discriminate particularly between animals and humans. For this reason, many of the predictions of the model for hippocampal-damaged animals are equally predicted in hippocampal-damaged humans. Of course, human amnesics typically have diffuse and varying degrees of damage; still, it may be possible to obtain similar behavioral profiles in these subjects on simple conditioning paradigms. We are currently conducting several experimental studies to test hippocampal-damaged amnesics on just these paradigms, using classical eyeblink conditioning and computer-based games which embed the logical structures of the conditioning paradigms in associative tasks. If the humans do behave as expected, these tasks may provide a way to assess residual hippocampal function, as well as granting insight into the similarities and differences between hippocampal function in humans and animals.

What New Modeling Studies Will This Work Lead To?

The studies described here are initial attempts to extend the original model in several ways; future work will continue these studies. Further work will be required to continue trying to instantiate the proposed top-down processes via bottom-up modules; particularly, it will be important to consider bottom-up models of other hippocampal region structures, the roles of other neuromodulators, and interaction with cortex. This last will provide a means to evaluate consolidation and retrograde amnesia *in* the model. Another important direction is to expand the applicability of the model to a wider range of behaviors, including especially the role of the hippocampus in "one-shot," declarative memories as well as incrementally acquired ones.

Meanwhile, there are still many data from the classical conditioning domain which the model does not yet account for. Chief

among these are real-time behaviors such as trace conditioning, conditioning with multiple interstimulus intervals, and sequential occasion-setting. To date, we have made some progress on this issue, by developing a real-time model of cerebellar processing which is consistent with current understanding of anatomical and physiological substrates (Gluck et al., 1994a); combination of this bottom-up cerebellar model with a hippocampal region network expanded to allow real-time processing may allow us to address some of these issues while maintaining the broad explanatory power of the original model.

REFERENCES

Alvarado M, Rudy J (1995) A comparison of kainic acid plus colchicine and ibotenic acid-induced hippocampal formation damage on four configural tasks in rats. Behav Neurosci 109:1052–1062.

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

Ambros-Ingerson J, Granger R, Lynch G (1990) Simulation of paleocortex performs hierarchical clustering. Science 247:1344–1348.

Berger T, Orr W (1983) Hippocampectomy selectivity disrupts discrimination reversal learning of the rabbit nictitating membrane response. Behav Brain Res 8:49–68.

Berry S, Thompson R (1979) Medial septal lesions retard classical conditioning of the nictitating membrane response in rabbits. Science 205:209–211.

Buhusi CV, Schmajuk NA (1996). Attention, configuration, and hippocampal function. Hippocampus 6:621–642.

Bunsey M, Eichenbaum H (1993) Critical role of the parahippocampal region for paired associate learning in rats. Behav Neurosci 107:740–747.

Cahusac P, Rolls E, Miyashita Y, Niki H (1993) Modification of the responses of hippocampal neurons in the monkey during the learning of a conditional spatial response task. Hippocampus 3(1):29–42.

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

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):331–339.

Eichenbaum H, Mathews P, Cohen NJ (1989) Further studies of hippocampal representation during odor discrimination learning. Behav Neurosci 103(6):1207–1216.

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

Eichenbaum H, Cohen NJ, Otto T, Wible C (1992b) Memory representation in the hippocampus: Functional domain and functional organization. In: Memory organization and locus of change (Squire LR, Lynch G, Weinberger NM, McGaugh JL, eds, pp 163–204. Oxford: Oxford University Press.

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):819–827.

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

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

Gluck M, Myers C, Goebel J (1994b) A computational perspective on dissociating hippocampal and entorhinal function (Response to Eichenbaum, et al.). Behav Brain Sci 17:478–479.

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

Gluck MA, Myers CE, Thompson RF (1994a) A Computational Model of the Cerebellum and Motor-Reflex Conditioning. In: Introduction to neural and electronic networks (Zornetzer S, Davis J, Lau C (eds), pp 91–98. San Diego: Academic Press, Inc.

Gluck M, Goren O, Myers C, Thompson R (1996/in preparation) A higher-order recurrent network model of the cerebellar substrates of response timing in motor-reflex conditioning. J Cogn Neurosci

Good M, Honey RC (1991) Conditioning and contextual retrieval in hippocampal rats. Behav Neurosci 105(4):499–509.

Gormezano I, Kehoe EJ, Marshall BS (1983) Twenty years of classical conditioning research with the rabbit. Prog Psychobiol Physiol Psychol 10:197–275.

Grossberg S (1976) Adaptive pattern classification and recoding: Part I. Biol Cybern 23:121–134.

Hasselmo M, Schnell E (1994) Feedback regulation of cholinergic modulation and hippocampal memory function. Proc World Congress Neural Networks 2:729–734.

Hasselmo M, Schnell E, Barkai E (1995) Dynamics of learning and recall at excitatory synapses and cholinergic modulation in hippocampal region CA3. J Neurosci 15(7 Pt 2):5249–5262.

Hasselmo ME, Wyble BP, Wallenstein GV (1996) Encoding and retrieval of episodic memories: Role of cholinergic and GABAergic modulation in the hippocampus. Hippocampus 6:693–708.

Hinton G (1989) Connectionist learning procedures. Artificial Intelligence 40:185–234.

Hirsh R (1974) The hippocampus and contextual retrieval of information from memory: a theory. Behav Biol 12:421–444.

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

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

Kaye H, Pearce J (1987) Hippocampal lesions attenuate latent inhibition and the decline of the orienting response in rats. Q J Exp Psychol 39B:107–125.

Kohonen T (1984) Self-organization and associative memory. New York: Springer-Verlag.

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

Lubow R (1973) Latent Inhibition. Psychol Bull 79:398–407.

Mackintosh N (1983) Conditioning and associative learning. Oxford: Oxford University Press.

Marr D (1971) Simple memory: a theory for archicortex. Proc R Soc Lond [Biol] 841(262)23–81.

McClelland JL, Goddard H (1996) Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. Hippocampus 6:654–665.

McClelland J, McNaughton B, O'Reilly R (1995) Why we have complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 102:418–457.

McNaughton B (1989) Neuronal mechanisms for spatial computation and information storage. In: Neural connections, mental computations (Nadel L, Cover L, Culicover P, Harnish R, eds), pp. 285–350. Cambridge, MA: MIT Press.

McNaughton B, Chen L, Markus E (1991) "Dead reckoning," landmark learning, and the sense of direction: a neurophysiological and computational hypothesis. J Cogn Neurosci 3(2):190–202.

Moore J, Goodell N, Solomon P (1986) Central cholinergic blockade by scopolamine and habituation, classical conditioning, and latent inhibition of the rabbit's nictitating membrane response. Physiol Psychol 4(3):395–399.

Morris R (1983) An attempt to dissociate "spatial-mapping" and "work-

ing-memory" theories of hippocampal function. In: Neurobiology of the hippocampus (Seifert W, ed), pp 405–432. London: Academic Press.
Murre J (1994) A model for amnesia, Unpublished manuscript, Medical Research Council—Applied Psychology Unit. Cambridge, UK.
Murre JMJ (1996) TraceLink: A model of amnesia and consolidation of memory. Hippocampus 6:675–684.
Myers C, Gluck M (1994) Context, conditioning and hippocampal representation. Behav Neurosci 108(5):835–847.
Myers C, Gluck M (1996) Cortico-hippocampal representations in simultaneous odor discrimination learning: a computational interpretation of Eichenbaum, Mathews & Cohen (1989). Behav Neurosci 110:1–22.
Myers C, Gluck M, Granger R (1995) Dissociation of hippocampal and entorhinal function in associative learning: a computational approach. Psychobiology 23(2):116–138.
Myers CE, Ermita B, Harris K, Hasselmo M, Solomon P, Gluck MA (1996) A computational model of the effects of septohippocampal disruption on classical eyeblink conditioning. Neurobiol Learning Memory 66:51–66.
Nadel L, Willner J (1980) Context and conditioning: a place for space. Physiol Psychol 8(2):218–228.
O'Keefe J (1979) A review of the hippocampal place cells. Prog Neurobiol 13:419–439.
O'Keefe J, Nadel L (1978) The hippocampus as a cognitive map. Oxford: Clarendon University Press.
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):111–119.
Pearce J, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev 87:532–552.
Penick S, Solomon R (1991) Hippocampus, context and conditioning. Behav Neurosci 105(5):611–617.
Phillips R, LeDoux J (1994) Lesions of the dorsal hippocampal formation interfere with background but not foreground contextual fear conditioning. Learning Memory 1:34–44.
Port R, Patterson M (1984) Fimbrial lesions and sensory preconditioning. Behav Neurosci 98:584–589.
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.
Rescorla R, Wagner A (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In: Classical conditioning II: current research and theory (Black A, Prokasy W, eds), pp 64–99. New York: Appleton-Century-Crofts.
Rudy J, Sutherland R (1989) The hippocampal formation is necessary for rats to learn and remember configural discriminations. Behav Brain Res 34:97–109.
Rumelhart D, Zipser D (1985) Feature discovery by competitive learning. Cogn Sci 9:75–112.
Rumelhart D, Hinton G, Williams R (1986) Learning internal representations by error propagation. In: Parallel distributed processing: explorations in the microstructure of cognition (Rumelhart D, McClelland J, eds), pp 318–362. Cambridge, MA: MIT Press.
Sakai K, Miyashita Y (1991) Neural organization for the long-term memory of paired associates. Nature 354:152–155.
Schmajuk N, DiCarlo J (1990) Backpropagation, classical conditioning and hippocampal function (Technical Report No. Northwestern University.
Schmajuk N, DiCarlo J (1992) Stimulus configuration, classical conditioning and hippocampal function. Psychol Rev 99:268–305.
Schmaltz L, Theios J (1972) Acquisition and extinction of a classically conditioned response in hippocampectomized rabbits (Oryctolagus cuniculus). J Comp Physiol Psychol 79:328–333.
Sears LL, Steinmetz JE (1990) Acquisition of classically conditioned-related activity in the hippocampus is affected by lesions of the cerebellar interpositus nucleus. Behav Neurosci 104:681–692.
Solomon P, 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 89:1192–1203.
Solomon P, Solomon S, Van der Schaaf E, Perry H (1983) Altered activity in the hippocampus is more detrimental to classical conditioning than removing the structure. Science 220:329–331.
Solomon P, Groccia-Ellison M, Flynn D, Mirak J, Edwards K, Dunehew A, Stanton M (1993) Disruption of human eyeblink conditioning after central cholinergic blockade with scopolamine. Behav Neurosci 107(2):271–279.
Sutherland N, Mackintosh N (1971) Mechanisms of animal discrimination learning. New York: Academic Press.
Taube J (1991) Space, the final hippocampal frontier? Hippocampus 1(3):247–249.
Thompson R (1986) The neurobiology of learning and memory. Science 233:941–947.
Thompson RF (1988) The neural basis of basic associative learning of discrete behavioral responses. Trends Neural Sci 11(4):152–155.
Thompson R (1990) Neural mechanisms of classical conditioning. Philos Trans R Soc Lond [Biol] 329:161–170.
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.
Weiner I (1990) Neural substrates of latent inhibition: the switching model. Psychol Bull 108(3):442–461.
Widrow B, Hoff M (1960) Adaptive switching circuits. Inst Radio Eng Western Electronic Show and Convention Record 4:96–104.
Woodruff-Pak DS (1993) Eyeblink classical conditioning in H.M.: delay and trace paradigms. Behav Neurosci 107(6):911–925.