Dissociation of hippocampal and entorhinal function in associative learning: A computational approach

CATHERINE E. MYERS and MARK A. GLUCK
Rutgers University, Newark, New Jersey

and

RICHARD GRANGER
University of California, Irvine, California

Unsupervised stimulus–stimulus redundancy compression, one component of Gluck and Myers's (1993) representational theory of the hippocampal-region function, could emerge from the anatomy and physiology of the entorhinal cortex. This hypothesis is suggested by a physiologically and anatomically realistic model of the entorhinal cortex derived from a similar model of the olfactory cortex previously proposed by Ambros-Ingerson, Granger, and Lynch (1990). To the extent that entorhinal function can survive damage strictly limited to the hippocampal formation (the H lesion), this has implications for interpreting the behavioral consequences of lesions which either do or do not spare overlying cortical areas. In particular, we expect that the H lesion should not interrupt stimulus–stimulus redundancy compression, thereby sparing conditioning behaviors, such as latent inhibition, which are eliminated by broader (H++) lesions to the hippocampal region. However, such other behaviors as the context sensitivity of latent inhibition and of learned responses are expected to be affected by the H lesion. These predictions are consistent with empirical data. The theory also leads to several novel predictions for behavioral comparisons of intact, H-lesioned, and H++-lesioned animals on tasks such as sensory preconditioning, compound preconditioning, and easy–hard transfer. A major theme of this paper is to illustrate how a bottom-up model of cortical processing can be integrated with a top-down model of hippocampal-region function to yield a more complete mapping from physiology to behavior.

Although there have been many attempts to characterize the functional role of the hippocampal region in learning and memory, most have treated the region as a single functional unit (e.g., Buzsaki, 1989; Eichenbaum, in press; Gluck & Myers, 1993; Hirsh, 1974; McNaughton & Nadel, 1990; Mishkin, 1982; Nadel, 1992; Squire, 1987). The hippocampal region, however, comprises several distinct structures, including hippocampus fields CA1 through CA4, dentate gyrus, subicular complex, and overlying areas, including the entorhinal cortex. Recent advances in experimental methods (Jarrard, 1989; Zola-Morgan, Squire, Rempel, Clower, & Amaral, 1992) have made it possible to selectively study these component structures. This has led to several surprising discoveries in which behaviors previously thought to be hippocampal dependent have instead been shown to depend primarily on cortices overlying the hippocampus (e.g., Jarrard, 1993; Zola-Morgan & Squire, 1992).

In light of these recent experimental results, researchers have begun to propose and test putative divisions of function among distinct hippocampal-region structures (Bunsey & Eichenbaum, 1993; Eichenbaum, Otto, & Cohen, 1994; Lynch & Granger, 1992; Squire, 1992). This paper describes one such effort: a preliminary mapping of a previous computational model of hippocampal-region function onto distinct structures of this region. Our primary goal here was to identify the functional role of the entorhinal cortex in elementary associative learning tasks.

The starting point for this differentiation of function is a previous connectionist model of the aggregate functional role of the hippocampal region in classical "Pavlovian" conditioning. Gluck and Myers (1993) have proposed that the hippocampal region is critical for the construction of new stimulus representations that facilitate learning. These new representations are assumed to

116

compress and differentiate information reflecting both stimulus–stimulus and stimulus-outcome regularities in the environment. A connectionist network model which performs these various representational changes accurately predicts a broad range of conditioning behaviors observed in intact animals that are altered or impaired after damage to the hippocampal region (Gluck & Myers, 1993; Myers & Gluck, 1994).

This model of hippocampal-region processing is illustrative of the class of "top-down" models in behavioral neuroscience which start with a comparison between the behavior of normal and lesioned subjects, and then try to derive underlying neural mechanisms than can account for the difference between these two populations (Gluck & Granger, 1993). Although such top-down models may not immediately yield detailed physiological understanding of the brain region in question, they do represent an important first step toward understanding the functional role of specific brain regions. Most importantly, this modeling approach ensures that the functional interpretation of a neural system is consistent with the broad base of detailed behavioral analyses provided by psychological studies of animal learning behavior.

A different, but complementary, approach to computational modeling in neuroscience begins not with behavior, but rather with the underlying anatomy and physiology, from which behavior might be derived as an emergent property of the neural circuitry. A major theme of this paper is to illustrate how such bottom-up modeling of cortical processing can be integrated with our existing top-down model of hippocampal-region function, to yield a more complete mapping from physiology to behavior.

We hypothesize that the anatomy and physiology of the superficial entorhinal cortex was consistent with an emergent clustering function. This suggestion parallels and draws on a recent bottom-up model of clustering in the anatomically similar superficial piriform cortex (Ambros-Ingerson, Granger, & Lynch, 1990; Granger, Ambros-Ingerson, & Lynch, 1989; Granger & Lynch, 1991); while the piriform cortex is limited to the olfactory domain, the entorhinal cortex could cluster information across the entire spectrum of sensory modalities (see also Gluck & Granger, 1993). This clustering is identical to the stimulus–stimulus compression mechanism proposed as one aspect of hippocampal region function in the top-down model (Gluck & Myers, 1993).

Our hypothesis suggests that the multiple components of Gluck and Myers's (1993) proposed hippocampal-region function could be dissociated and localized in the various structures of the region. To the extent that these subfunctions can exist in isolation, hippocampal lesion would be expected to yield different behavioral effects depending on the exact extent of the lesion. There is now behavioral evidence confirming that at least some entorhinal processing can survive such a localized hippocampal lesion (Jarrard & Davidson, 1991; Zola-Morgan & Squire, 1993).

This paper reviews Gluck and Myers's (1993) top-down theory of corticohippocampal processing and the bottom-up piriform cortex model of Ambros-Ingerson et al. (1990), and then argues that their model can also be interpreted as a bottom-up model of superficial entorhinal cortex. This entorhinal model can be included in a system that represents processing after a selective hippocampal lesion; the behavior of this system is compared with existing behavioral data on classical conditioning following such selective lesions. The model correctly predicts that selective hippocampal lesion should spare latent inhibition but not the context sensitivity of learned responses or of latent inhibition. The model also makes several novel (and still untested) predictions that behaviors should be spared or eliminated after the selective lesion, depending on the extent to which they depend on stimulus–stimulus redundancy compression. Finally, we compare and contrast this hypothesis with other related models of entorhinal and hippocampal function.

A "TOP-DOWN" MODEL OF
HIPPOCAMPAL-REGION FUNCTION

Early psychological theories of learning sought to characterize the nature of what an animal learns in a conditioning paradigm by formalizing rules for the development of elementary associations (Hull, 1943; Pavlov, 1927). These experimental studies led to the development of many one-process models of conditioning which described the formal properties of changes in underlying associations (Estes, 1958; Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972). Later behavioral studies, however, suggested that it was not possible to reduce to a single associative weight all that an animal learns during a conditioning experiment. Phenomena which most clearly implicate the need for multiple learning mechanisms include latent inhibition (Lubow, 1973), sensory preconditioning (Thompson, 1972), and reversal transfers and shifts (Kendler, Kendler, & Sanders, 1967; Lawrence & Mason, 1955). These and other similar studies led to the development of two-process theories of learning in which the relevance of a stimulus cue could be altered independently from changes in associative weights (Frey & Sears, 1978; Schmajuk & Moore, 1988; Trabasso & Bower, 1964; Wagner, 1978). These two-process models allow for both *stimulus selection*, in which the associability, attention, or representation of a stimulus cue is selectively altered, as well as for *stimulus association*, in which the selected aspects of the stimulus event are associated with various other stimuli, especially reinforcing outcomes.

Building on the psychological tradition and behavioral data which support the distinction between stimulus selection and stimulus association, Gluck and Myers (1993) proposed a computational theory of hippocampal function in classical conditioning in which the hippocampal region is presumed to play a critical role in mediating stimulus selection but is not necessary for the formation of elementary stimulus associations. Central to Gluck and Myers's computational account of stimulus selection and hippocampal-region function is the concept of a modifi-

able *stimulus representation*. Because the similarity between two stimulus representations determines the extent to which associations learned to one stimulus will generalize to the other, changes in the underlying stimulus representations can alter the amount of generalization between stimuli. Thus, if two stimuli are mapped to the same response, learning will be facilitated if their representations are made more similar, or *compressed*, to increase generalization between them. Conversely, representations can be *differentiated*, or made less similar, to decrease generalization if the stimuli are to be mapped to different responses. The hippocampal region is assumed to compress and differentiate stimulus representations in exactly this way.

In Gluck and Myers's (1993) theory, two constraints are assumed to govern the degree of compression and differentiation within the hippocampal-region representations. First, *stimulus–stimulus* regularities, that is, the reliability with which two stimuli co-occur, will tend to cause representational compression. This follows from noting that if two stimuli are perfectly correlated, generalization between them should be increased. The second constraint depends on *stimulus-outcome* regularities: two stimuli which reliably predict the same future outcome (a salient or reinforcing event) should have their representations compressed, while stimuli which predict different events should have differentiated representations. Table 1 summarizes these conjectured constraints on the formation of new stimulus representations in the hippocampal region.

The biases to compress and differentiate representations may often be at odds with each other. For example, if stimulus A always occurs in context X, there will be a bias to compress the representations of A and X. If X often occurs without A, there will be a bias to differentiate their representations. Which bias prevails will depend largely on the statistical correlations governing the appearance of each cue and their mutual predictiveness. The stimulus–stimulus and stimulus-outcome processes may also be at odds. If two stimuli never co-occur, stimulus–stimulus processes will tend to differentiate their representations. If both stimuli reliably predict the same salient future event, stimulus-outcome processes will tend to compress their representations. Which process dominates will de-

Table 1
Representational Changes Assumed to Occur in the Hippocampal Region During Learning

	Bias to Compress	Bias to Differentiate
Stimulus–stimulus regularity	IF stimuli reliably co-occur	IF stimuli reliably occur separately
Stimulus-outcome regularity	IF stimuli reliably predict same future event	IF stimuli reliably predict different future events

Note—The hippocampal region is assumed by Gluck and Myers (1993) to compress and differentiate stimulus representations to increase and decrease stimulus generalization. This process is constrained by stimulus–stimulus regularity—how reliably stimuli co-occur—as well as by stimulus-outcome regularity—how reliably stimuli predict the same future events.

pend on the strength of the statistical correlations, as well as on number of exposures. Early in training, before stimulus-outcome regularities are detected, stimulus–stimulus processes may tend to dominate, inasmuch as stimulus–stimulus correlations can be estimated after only a few exposures. If, however, stimuli predict salient outcomes, such as the presence or absence of reinforcement, stimulus-outcome processes may eventually dominate the representation.

A system that reflects these representational processes has been implemented as a connectionist model (Gluck & Myers, 1993; Myers & Gluck, 1994), as shown in Figure 1A. One network, a predictive autoencoder (Hinton, 1989), represents processing in the hippocampal region. This network is trained by error backpropagation (Parker, 1985; Rumelhart, Hinton, & Williams, 1986; Werbos, 1974); it learns to map from inputs representing stimulus activity, through a narrow internal or hidden layer of nodes, to output nodes which learn to reconstruct the stimulus inputs plus a prediction of future events. Because the internal layer is narrow, as compared with the input and output layers, the representations formed in the internal layer must compress redundant information while preserving enough information to allow the output to be reconstructed. In this process, the representations tend to compress and differentiate information on the basis of both stimulus–stimulus and stimulus-output regularities, just as the hippocampal region is proposed to do. It is important to note that this predictive autoencoder, with its training algorithm, represents only one possible implementation of the proposed hippocampal-region function; later in this paper we discuss how part of the function may be implemented in such a way as to map more directly onto the biological substrate.

The model further assumes that extrahippocampal regions in the cerebral and cerebellar cortices are the sites of long-term memory. These sites are assumed to be capable of simple stimulus association, but not of forming new stimulus representations. One such long-term memory network, which is shown on the left in Figure 1A, is a simplification of some aspects of motor-reflex conditioning in the cerebellum (Gluck, Goren, Myers, & Thompson, 1994). This network has access to a simple correlational learning rule, the Widrow–Hoff rule (Widrow & Hoff, 1960), which is related both to psychological descriptions of learning (Gluck & Bower, 1988; Rescorla & Wagner, 1972; Sutton & Barto, 1981) and to biological plasticity mechanisms such as long-term potentiation (LTP) (Bliss & Lomo, 1973; Levy, Brassel, & Moore, 1983; Stanton & Sejnowski, 1989). This learning rule is sufficient to allow the network to learn to map from its existing representations to an output that is interpreted as the behavioral response, but not to develop new representations. However, a separate application of this learning rule can allow the network to adopt the new representations constructed in the internal layer of the hippocampal-region network. Figure 1A shows sparse, fixed connections between the internal layers of the two networks, allowing the long-term memory network to copy (a re-

(A)

Long-Term Memory   **Hippocampal Region**

Response

Reinforcement

Stimulus Input

Training signal

Stimulus Input

(B)

Long-Term Memory
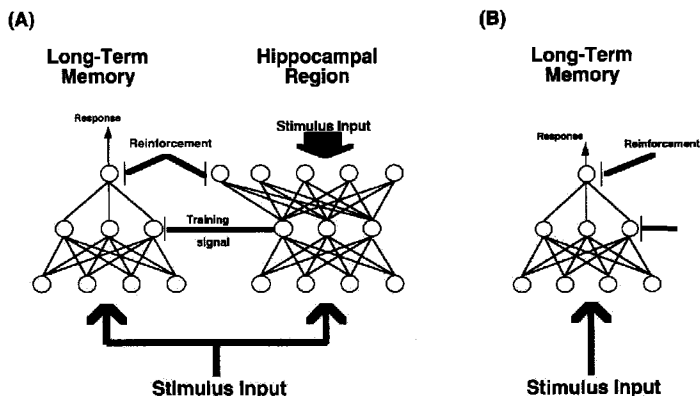
Response   Reinforcement

Stimulus Input

Figure 1. A connectionist implementation of the top-down hippocampal-region model of Gluck and Myers (1993). (A) The intact model: Stimulus input is presented to the hippocampal-region network, which can compress and differentiate stimulus representations. These representational changes are biased by stimulus–stimulus redundancy as well as by stimulus–outcome predictiveness. The new representations constructed in the hippocampal-region network are acquired by the long-term memory network, which simultaneously learns to map from these internal representations to an output interpreted as the system's behavioral response. This long-term-memory network is based on a prior model of some aspects of motor-reflex learning in the cerebellum (Gluck et al., 1994). (B) The H^{++}-lesioned model: Broad hippocampal-region damage (the H^{++} lesion) is simulated by disabling the hippocampal-region network. In this case, the remaining long-term-memory network cannot form new internal representations, although it can still learn new mappings to behavioral responses based on its preexisting (and now fixed) internal representations.

coding of) the evolving hippocampal-region representations. Learning in the hippocampal-region network and in both layers of the long-term memory network is assumed to proceed incrementally and in parallel. This model does not focus on the relative speeds of learning in the two networks or on the time course of transfer of information from one network to the other; however, this general two-component architecture to learning would certainly allow for a consolidation period during which information is slowly transferred from hippocampal region to long-term memory (see Alvarez & Squire, 1994; McClelland, McNaughton, & O'Reilly, 1994; Murre, 1994; Squire, 1987; Winocur, 1990).

Within this connectionist model, broad hippocampal-region damage (often termed the H^{++} lesion) is simulated by disabling the hippocampal-region network (Figure 1B). In this case, the long-term memory network can still learn behavioral responses on the basis of its existing representations, but can no longer modify these representations during learning. Our H^{++}-lesioned model therefore predicts that hippocampal-region damage will be most deleterious to those tasks which require new stimulus representations, but not those for which preexisting or random stimulus representations suffice.

For example, in rabbit eyeblink conditioning, a blink-evoking corneal airpuff or paraorbital shock (the unconditioned stimulus, or US) is repeatedly preceded by a neutral cue, such as a tone or light (the conditioned stimulus, or CS); over time, an association develops between

CS and US such that the CS alone can come to elicit an anticipatory blink, the conditioned response, or CR (Gormezano, Kehoe, & Marshall, 1983). Under optimal conditions, the acquisition of a CR in this preparation is not affected by hippocampal lesion (Port, Mikhail, & Patterson, 1985), consistent with the model's prediction that elementary associative learning should not depend on the hippocampal region. However, more complex paradigms which involve significant representational changes are substantially altered by damage to the hippocampal region. The intact and H++ lesioned models have been applied to a variety of these paradigms (Gluck & Myers, 1993; Myers & Gluck, 1994), and we describe a few examples below.

Behavioral Results: Compression and Differentiation

One constraint assumed to bias hippocampal-region representations is stimulus–stimulus compression of co-occurring stimuli. For example, if A and B are repeatedly presented together, their representations become compressed, increasing generalization between them. As a result, subsequent training to A will transfer to B as well. The response to B alone will be greater under these conditions than in a control condition in which there is no preexposure, as shown in Figure 2A (Gluck & Myers, 1993). This phenomenon, known as *sensory preconditioning*, occurs in intact animals as well (Thompson, 1972). The H++-lesioned model, with no representational com-
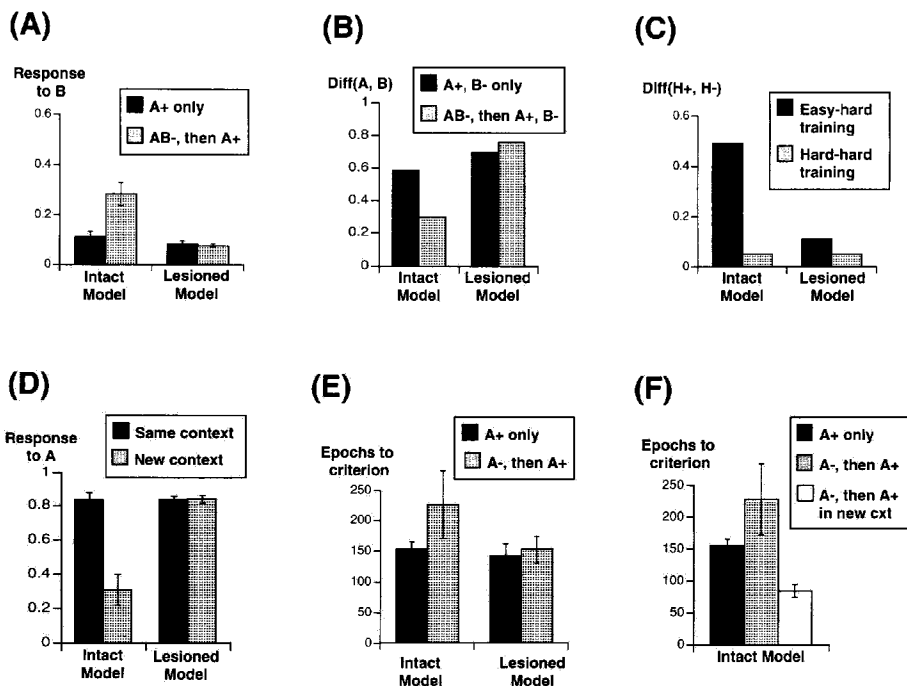
(A)

(B)

(C)

(D)

(E)

(F)

Figure 2. Simulation results with the intact and H^{++}-lesioned models, replotted from results reported in Gluck and Myers (1993) and Myers and Gluck (1994). (A) Sensory preconditioning: Unreinforced preexposure to a compound of two stimuli, AB, followed by training to respond to A, results in stronger responding to B alone than in a control condition with no preexposure, in the intact but not the H^{++}-lesioned model. Fornix lesion similarly eliminates sensory preconditioning in rabbits (Port & Patterson, 1984). (B) Compound preconditioning: Unreinforced preexposure to AB slows later training to discriminate A and B, as shown by the less relative difference in responding to A and B [Diff(A,B)] in the intact model. Intact rats show compound preconditioning (Lubow et al., 1976); the model predicts H^{++} lesion should eliminate the effect. (C) Easy–hard transfer: Prior training on an "easy" discrimination speeds later learning of a "hard" discrimination, as shown by the greater relative difference in responding to the hard stimuli [Diff($H+,H-$)], in the intact model. Normal animals show this effect (e.g., Lawrence, 1952); the model predicts that H^{++} lesion should eliminate the effect. (D) After training to respond to a cue A in some context, the intact, but not the H^{++}-lesioned, system shows a decrement in responding to A when presented in a novel context. Hippocampal lesion does eliminate the effect in rabbits (Penick & Solomon, 1991). (E) Latent inhibition: Unreinforced preexposure to a cue, A, slows later training to respond to A in the intact, but not the H^{++}-lesioned, model. Hippocampal aspiration or electrolytic lesion likewise eliminates latent inhibition in animals (Kaye & Pearce, 1987; Solomon & Moore, 1975). (F) The intact system shows release from latent inhibition with a context shift between preexposure and training phases. Likewise, latent inhibition is eliminated in intact rats (Bouton & Brooks, 1993; Lubow et al., 1976) by such a shift.

pression, does not show any sensory preconditioning effect; fimbrial lesion, which disrupts hippocampal processing, is likewise sufficient to eliminate sensory preconditioning in the rabbit eyeblink preparation (Port & Patterson, 1984).

Unreinforced preexposure to the compound AB retards later learning to discriminate A and B in the intact model, for reasons similar to those described above for sensory preconditioning; compression during the preexposure phase increases generalization between A and B, making it harder to subsequently map them to different responses (Gluck & Myers, 1993). Figure 2B shows this *compound preconditioning* effect, which is also ob-

served in intact rats and normal children (Lubow, Rifkin, & Alek, 1976); the relative difference Diff(A,B) in response rates to the two stimuli following 100 training epochs is reduced after 20 AB preexposure trials, relative to a control condition with no preexposure. Again, this effect is eliminated in the H^{++}-lesioned model; to our knowledge, this is a novel prediction which remains to be tested in lesioned animals.

Other behavioral results are assumed to result from stimulus-outcome differentiation during learning. One such phenomenon is *easy–hard transfer*, in which early training on an "easy" discrimination (e.g., black vs. white) facilitates subsequent learning of a "hard" discrimina-

tion along the same dimension (e.g., dark gray vs. light gray) more than does an equivalent amount of pretraining on the hard task itself. During training on the easy task, the hippocampal-region network is assumed to differentiate the representations of the two easy stimuli, which predict different outcomes. This differentiation will result in decreased generalization among all stimuli lying along the same stimulus dimensions. As a result, subsequent learning of the hard task is facilitated (Gluck & Myers, 1993). Figure 2C shows that the relative difference Diff(H+,H−) in response rates to the two stimuli in the hard task is greater following pretraining on the easy task than after equal pretraining on the hard task itself. This effect is also obtained in intact animals (e.g., Lawrence, 1952; Riley, 1968; Terrace, 1963), even if the hard task is a reversal of the easy task (Mackintosh & Little, 1970). With no mechanisms for stimulus-outcome differentiation, the H++-lesioned model does not show the easy–hard transfer effect. This prediction that easy–hard transfer should depend on the hippocampal region has not been tested in hippocampal-lesioned animals and remains a novel prediction of the theory.

Behavioral Results: Contextual Processing

This account of hippocampal-region function also provides an interpretation for many contextual effects in intact and lesioned animals. For example, under many conditions, intact animals trained to respond to a stimulus, A, in one context, X, show a decreased responding if A is presented in a new context, Y (Antelman & Brown, 1972; Good & Honey, 1991; Honey, Willis, & Hall, 1990; Kim & Fanselow, 1992; Penick & Solomon, 1991). Figure 2D shows this response-decrement effect in the intact model as well (Myers & Gluck, 1994). During training to respond to A in context X, several representational processes operate in the intact model. Since A always appears in conjunction with X, there is a bias toward stimulus–stimulus compression; this is opposed by stimulus–stimulus differentiation which operates because the context X often occurs without A. At the same time, the system learns that reinforcement is contingent on the presence of A—but not X alone—and so stimulus-outcome mechanisms will tend to differentiate the representations. These various constraints are summarized in Table 2.

Which of these various representational mechanisms dominates depends on environmental regularities and on the amount of training. Early in training, before the correlation between A and reinforcement is learned, stimulus-outcome mechanisms will be weak. On the other hand, stimulus–stimulus regularities can be determined after only a few exposures. Since A perfectly predicts X but X predicts either A's presence or A's absence, there will be a net tendency for stimulus–stimulus compression of A and X. If A is subsequently presented in a new context, Y, the new inputs will activate this representation of A less strongly than usual; in turn, the behavioral response may be less strongly activated than usual, as shown in Figure 2D. The H++-lesioned model does not show this

Table 2
Representational Changes During Response Acquisition

	Bias to Compress	Bias to Differentiate
Stimulus–stimulus regularity	$A \Rightarrow X$	$X \Rightarrow A$, $X \Rightarrow$ not A
Stimulus-outcome regularity		$AX \Rightarrow US$, X alone $\Rightarrow$ no US

Note—Representational changes during training to respond to cue A in context X. A appears only in context X, and so there is a bias to compress the representations of A and X. Since X often appears without A, some stimulus–stimulus differentiation also occurs. Since A predicts reinforcement (the US) while X alone does not, there is stimulus-outcome differentiation of A and X. The net effect is to differentiate the representations of A and X, decreasing generalization between them, and facilitating learning to discriminate them. However, because stimulus-outcome regularities may become apparent only after some training, stimulus–stimulus compression may dominate early in training.

effect, since it has no compression mechanisms; consistent with this expectation, ablation of dorsal hippocampus and subiculum suffices to eliminate the context-shift decrement in rabbits (Penick & Solomon, 1991).

However, the representations in the intact model continue to evolve with time. As the mapping between A and US is repeatedly trained, stimulus-outcome processes eventually come to dominate, decreasing generalization between A and X, to facilitate learning to respond only when A is present. As the representation of A becomes highly differentiated from X, the response to A is activated strongly, regardless of the context, and remains strong even in the new context, Y. The response decrement shown in Figure 2D for the intact model is only a transient phenomenon and is eliminated after overtraining the intact model in the original context (see Myers & Gluck, 1994, for further details). This is consistent with data from the conditioned emotional response paradigm showing that the response decrement after context shift is eliminated with extensive training in the original context (Hall & Honey, 1990).

Another context-related behavioral phenomenon is *latent inhibition*: unreinforced preexposure to cue A slows later acquisition of a response to A (Lubow, 1973). The intact model shows this effect as well (Figure 2E; see also Myers & Gluck, 1994). During the preexposure phase, there are opposing biases to compress and differentiate the representations of A and the context, X, in which it occurs, by virtue of the fact that A perfectly predicts X, but X only partially predicts A. There is also a very strong tendency toward stimulus-outcome compression, since both A and X are equally predictive of no salient outcome. Table 3 summarizes these various representational constraints. The net result is therefore a tendency to compress the representations of A and X. This compression increases the generalization between A and X, and also simply decreases the amount of representational resources allocated to encode specifically for the presence of A. Both of these processes interfere with subsequent learning to respond to A but not to X alone; stimulus-outcome differentiation comes to undo this compression, allowing the discrimination to be acquired, but learning

Table 3
Representational Changes During Unreinforced Preexposure

	Bias to Compress	Bias to Differentiate
Stimulus–stimulus regularity	$A \Rightarrow X$	$X \Rightarrow A$, $X \Rightarrow$ not A
Stimulus–outcome regularity	$AX \Rightarrow$ no US, X alone $\Rightarrow$ no US	$AX \Rightarrow$ no US, X alone $\Rightarrow$ no US

Note.—Representational changes during preexposure to cue A in context X. A appears only in context X, and so there is a bias to compress the representations of A and X. Since X often appears without A, some stimulus–stimulus differentiation also occurs. Since neither A nor X predict reinforcement (the US), there is a strong tendency for stimulus-outcome compression of A and X. The net effect is to compress the representations of A and X, increasing generalization between them and impeding subsequent learning to discriminate them.

is slower than in a control condition where there was no initial compression of cue and context. Again, the H^{-+}-lesioned model does not show this effect, consistent with data showing that latent inhibition is also eliminated after hippocampal aspiration or electrolytic lesion (Kaye & Pearce, 1987; Solomon & Moore, 1975; but see also below).

Since latent inhibition is assumed to result from increased generalization between cue A and context X, and a resulting decrease in representational resources allocated to encode for A, a context shift between preexposure and acquisition phases interferes with the effect. There is immediate release from the problem of overgeneralization between A and X—at least if the new context Y is sufficiently different from X. While there is still the problem that the representation of A has been reduced, that can be overcome relatively quickly by strong stimulus-outcome differentiation mechanisms in the subsequent acquisition phase. As a result, latent inhibition is eliminated in the intact model by a context shift between preexposure and acquisition phases (Myers & Gluck, 1994). Figure 2F shows that the context shift may even speed learning, relative to a control condition: if the acquisition in context Y is very different from the preexposure context X, and the representation of A has been compressed with context X, the net result may be a representation which already distinguishes A from Y. Similarly, in intact rats and in children, preexposure to a stimulus in one context has been found to facilitate subsequent learning about that stimulus in a novel context (Lubow et al., 1976), reversing the latent inhibition effect found when the stimulus is trained in the original context. Context switch also eliminates latent inhibition in a rat conditioned-suppression paradigm (Bouton & Brooks, 1993).

MODELING THE SELECTIVE CONTRIBUTIONS OF ENTORHINAL CORTEX TO HIPPOCAMPAL-REGION PROCESSING

The model behaviors described above arise from the interplay, and often opposition, between representational compression and differentiation and between stimulus–

stimulus and stimulus-outcome correlations. Stimulus–stimulus learning may be quite fast, while stimulus-outcome learning, which requires detecting stimulus-reinforcement correlations, may be slower but stronger in the end. Given that these processes can be partially dissociated and that they occur at different speeds and strengths, it may be possible that they represent different subfunctions of the hippocampal region and that they are situated in different anatomical loci.

The top-down hippocampal-region model presented by Gluck and Myers (1993), and described above, considers these processes to operate within an undifferentiated system representing the aggregate representational role of the hippocampal region. The hippocampal region actually consists of many highly differentiated structures, as shown in Figure 3. The primary pathway for sensory input to the hippocampal region is through the superficial entorhinal cortex, which, in turn, is afferented by the perirhinal cortex, the parahippocampal cortex, and a full range of higher sensory and associational cortices (Schousboe et al., 1993; van Hoesen & Pandya, 1975); from the entorhinal cortex, information travels through the hippocampal formation, including the dentate gyrus and fields CA1 through CA4 of the hippocampus before returning through deep layers of entorhi-
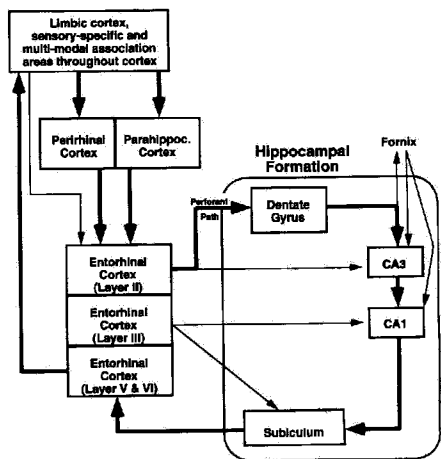

Figure 3. Schematic representation of major information flow pathways in the hippocampal region. The entorhinal cortex receives signals from the limbic cortex and sensory-specific and multimodal association areas (via the perirhinal and parahippocampal cortices). The perforant pathway carries projections from the superficial entorhinal cortex to the dentate gyrus, and information travels thence to hippocampal fields CA3 and CA1, the subiculum, and back to the deep layers of the entorhinal cortex, which project back to the sensory association areas. There are many additional pathways not shown here, including connections between the fornix and subcortical structures, including the thalamus and septum. After Amaral, Ishizuka, & Claiborne (1990) and van Hoesen (1982).

nal cortex to the same cortical areas from which it originally arrived (Witter, Groenewegen, Silva, & Lohman, 1989). Various other pathways exist as well, particularly the fornix, which carries information bidirectionally between the hippocampus and such subcortical structures as the septum and thalamus (Swanson, 1979). An important consideration of any hippocampal-region model is how it can be mapped onto the local architecture and information pathways of the region.

As a first step towards this goal, we consider here what might be the particular contribution of the superficial entorhinal cortex to hippocampal-region processing. We consider a model which—although originally proposed as a model of piriform cortex—also captures many aspects of entorhinal anatomy, physiology, and plasticity. This model performs stimulus–stimulus clustering of its inputs, related to the redundancy compression which is assumed to constrain hippocampal-region representations (Gluck & Granger, 1993). Thus, we propose that, in the same way as the piriform cortex may cluster olfactory information, the entorhinal cortex may cluster information across the full polymodal range of its inputs. In terms of our larger theory of hippocampal-region function, we propose that representational functions may be performed in different subregions, with the entorhinal cortex contributing a stimulus–stimulus redundancy compression function and other structures adding further representational constraints, possibly in a largely sequential fashion.

A Model of Learning in Piriform Cortex

Granger and his colleagues (Granger & Lynch, 1991; Granger et al., 1989; Ambros-Ingerson et al., 1990) have noted that hierarchical clustering of odor inputs can emerge from the superficial layers of the primary olfactory (piriform) cortex. A computational model of this process, based on the induction and expression rules for synaptic strengthening via LTP in the piriform cortex (Jung, Larson, & Lynch, 1990; Kanter & Haberly, 1990), is shown in Figure 4. In brief, LTP of bulb–cortex synapses in the model causes an increase in the similarity of cortical responses to any of a range of similar inputs. The cortical response thereby comes to correspond to a family or cluster of inputs: a given cortical response simply signals membership of the input in a given cluster (e.g., "fruit" odors vs. "meat" odors vs. "floral" odors). This functional operation is equivalent to statistical clustering, and has been previously described as a consequence of correlative learning rules in network models by many researchers (Grossberg, 1976; Kohonen, 1984; Rumelhart & Zipser, 1985; von der Malsburg, 1973).

This clustering phenomenon emerges solely from feedforward activity in the olfactory system model. Incorporation of the extensive feedback system from cortex back to the olfactory bulb causes iterative responses from the bulb–cortex system from which more complex structure emerges; in particular, finer grained subclusters evolve within the initial coarse clusters. After having learned a number of floral odors, the initial sampling of another
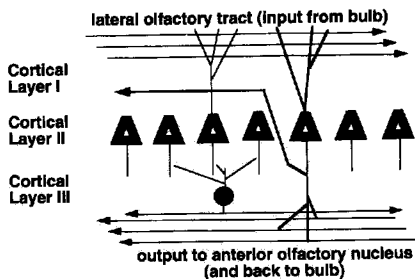

Figure 4. Schematic representation of olfactory system model proposed by Ambros-Ingerson et al. (1990), including feedforward connections from the olfactory bulb to the piriform cortex, and feedback inhibition from the cortex to the bulb. Layer II pyramidals are inhibited by layer III interneurons with dense local connectivity; layer II activity in turn forms an inhibitory mask on bulb activity. A model of these systems, involving a learning rule based on physiological mechanisms of repetitive sampling (the olfactory–hippocampal theta rhythm) and correlative characteristics of synaptic long-term potentiation, learns to self-organize a similarity-based hierarchy for classifying stimuli.

floral odor (e.g., a rose) causes feedforward activation of those cortical cells associated with the floral cluster. Learning in the feedback pathway from these cortical cells back to the bulb inhibitory layer causes that feedback inhibition to be selective to those portions of the bulb response correlated with the cortical "floral" cluster response. Resampling of the odor arrives against the background of this feedback inhibition, masking out the portion of the bulb response that corresponds to the floral cluster and passing on only the portion of the response specific to the particular (rose) odor. The resulting new cortical response therefore corresponds to those components of the odor that differ among various members of the cluster. Subsequent samples yield ever finer grained subclusters, distinguishing individual odors. Thus, the model generates a multilevel hierarchical memory which is traversed sequentially from general to more specific during recognition (Ambros-Ingerson et al., 1990).

Applying the Piriform Cortex Model to the Entorhinal Cortex

The posterior extent of the piriform cortex elides continuously with the adjacent lateral entorhinal cortex, and their superficial layers are closely related anatomically and physiologically, suggesting the possibility of related functionality (Price, 1973; van Hoesen & Pandya, 1975; Woodhams, Celio, Ulfig, & Witter, 1993). Noting this anatomical similarity, Gluck and Granger (1993) proposed that the entorhinal cortex may perform functions analogous to those identified in the primary olfactory cortex: similarity-based clustering of inputs. The relevance of the olfactory modeling studies to the entorhinal cortex is based on anatomical, physiological, and plasticity characteristics of the two regions.

Anatomy. The primary anatomical characteristics modeled in the superficial layers of the piriform cortex were apical dendrites, a ratio of roughly 100:1 excitatory-to-inhibitory cells, radial axonal arborization of inhibitory interneurons sufficient to contact roughly 100 excitatory cells, and random feedforward synaptic connectivity. All these feedforward constraints hold for the superficial entorhinal cortex as well. The superficial entorhinal cortex contains small layer II pyramidal or modified stellate cells with spiny apical dendrites which receive sparse nontopographic synaptic contact from afferents in layer I (van Hoesen & Pandya, 1975). Both AMPA- and NMDA-type glutamate receptors are apparently colocalized at the synaptic targets of these afferents (Monaghan & Cotman, 1985). As in the olfactory system, the primary layer II cells receive feedforward and feedback inhibition from GABAa and GABAb cells in layers I and II (Kohler, Wu, & Chan-Palay, 1985), and the feedback interneurons densely contact neighboring excitatory cells within a limited radius (Kohler, 1986). Lacking feedback connections to inhibitory cells in its inputs structures, the entorhinal cortex would not be expected to perform the repetitive cyclic subclustering found in the piriform model, but the feedforward clustering step is supported by the similar anatomy.

Physiology. The primary physiology modeled in the olfactory system study included the well-established synchronous rhythmic firing of the system at the 5- to 8-Hz theta rhythm during exploration and learning in freely moving animals (Macrides, 1975; Macrides, Eichenbaum, & Forbes, 1982), dendritic summation approximated by linear sums with superposition (Anton, Granger, & Lynch, 1992, 1993; Anton, Lynch, & Granger, 1991), lateral excitatory-inhibitory interactions over the period of individual theta peaks causing competitive behavior (Coultrip, Granger, & Lynch, 1992). All of these modeled features are equally applicable to the superficial entorhinal network.

Plasticity. Synaptic LTP has been shown to occur in piriform cortex, induced via brief bursts of activity at the naturally occurring theta rhythm, both in vivo (Roman, Staubli, & Lynch, 1987) and in vitro (Jung et al., 1990; Kanter & Haberly, 1990). This form of synaptic plasticity possesses the key properties required of a learning mechanism—rapid induction (seconds), long duration (weeks), and high capacity (via synapse specificity)—and is thus considered a leading candidate mechanism for learning (see Granger & Lynch, 1991, for a review). LTP with the same characteristics (NMDA-dependent, theta-induced) has been reported in the superficial entorhinal cortex (de Curtis & Llinas, 1993).

In sum, there are important points of correspondence between the anatomical, physiological, and plasticity characteristics of superficial entorhinal and piriform cortices, including the primary characteristics that gave rise to the emergent functionality found in the modeling studies of the piriform cortex. To the extent that entorhinal and piriform cortices share these properties, similar functionality may be expected to emerge.

A notable difference between these two cortical structures is their input. While the piriform cortex receives afferents only from the olfactory bulb, the entorhinal cortex receives inputs from the full spectrum of modalities as well as from supramodal association cortices and paralimbic areas (Woodhams et al., 1993). Thus, while the piriform cortex might perform hierarchical clustering of olfactory signals, the entorhinal cortex could perform clustering among stimuli in different modalities (e.g., sound and vision) or among the polymodal features of a single stimulus.

A second important difference, already noted, between the piriform and entorhinal cortices is the lack of inhibitory feedback connections from the entorhinal cortex to its inputs, which gave rise to masking of portions of the input in the olfactory system model. In the absence of such feedback, only the initial feedforward step of nonhierarchical clustering would be expected to occur. There are intense connections from the entorhinal cortex back to the neocortex (van Hoesen, 1982), but instead of performing a masking function, our model assumes that these connections guide the storage of hippocampal-region-mediated representations in neocortical long-term storage sites.

The hypothesized model of entorhinal function may thus be summarized as follows (full implementation details are given in the Appendix). The network consists of units representing superficial entorhinal excitatory layer II cells, sparsely afferented by multimodal inputs. The units are grouped into patches whose members are assumed to contact and be contacted by an inhibitory interneuron. The result of excitatory-inhibitory interaction in each patch is the emergence of lateral competition, which modeling has shown causes only the most activated targets to respond, thereby approximating "winner-take-all" activity. The activation of these winning units across the network defines a new representation of the stimulus. The winning nodes undergo LTP-like increment to increase their likelihood of winning the competition when similar inputs are presented in the future. This network, shown in Figure 5, performs unsupervised learning, since the clusters formed are independent of any reinforcement or stimulus-outcome pairing relationships. Its operation is similar to many other unsupervised competitive clustering networks proposed previously (e.g., Grossberg, 1976; Kohonen, 1984; von der Malsburg, 1973).

Relationship to Postulated Hippocampal-Region Function

In the previous section, an analogy between models of piriform and entorhinal cortex suggested that the superficial entorhinal cortex could perform an unsupervised clustering of its inputs. This clustering would reflect surface similarities among stimuli, but it would also reflect stimulus-stimulus redundancy. If two stimulus features appear together, they will be treated by the model as a single, compound stimulus, and will be assigned membership in a cluster according to the most salient features of each component stimulus. If one of the component fea-
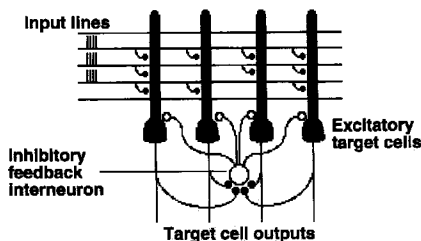
Figure 5. A model of the superficial entorhinal cortex, adapted from the piriform cortex model shown in Figure 4. Stimulus inputs sparsely excite target pyramidal cells, which in turn activate local feedback inhibitory cells. The feedback inhibition silences all but the most strongly activated pyramidal cells; these "winning" cells undergo synaptic plasticity, making them more likely to respond strongly to similar inputs in the future. Other "losing" cells undergo plasticity, making them less likely to respond to similar inputs. As emergent behavior, this unsupervised, competitive algorithm generates stimulus clustering: inputs with similar surface structure tend to be mapped to the same outputs. Additionally, co-occurring stimuli tend to be clustered. Additional input-masking processes in the olfactory system model, which depend on repetitive sampling mechanisms and on feedback connections between the piriform cortex and olfactory bulb, are not assumed in the entorhinal model.

tures is later presented alone, it will be treated as a degraded version of the compound, and will generally be assigned to the same cluster. Therefore, learning about the compound will generalize to the components (Lynch, 1986). This assignation of co-occurring stimuli to the same representational cluster effects the kind of stimulus–stimulus redundancy compression which Gluck and Myers (1993) previously proposed as one constraint biasing new stimulus representations in the hippocampal region.

We therefore hypothesize that the entorhinal cortex contains sufficient circuitry to implement stimulus–stimulus redundancy compression. This, in turn, implies that the remaining representational constraints shown in Table 1 would be implemented elsewhere in the hippocampal region. One possibility would be that various hippocampal-region structures each impose one kind of representational constraint and operate in assembly-line fashion, each contributing to the evolving representation. Such an assembly-line hypothesis has previously been proposed as consistent with the curious unidirectionality of connections within the hippocampal formation (Lynch & Granger, 1992): for example, the superficial entorhinal cortex generates many more afferents to the dentate gyrus and hippocampal field CA3 than it receives back from them (Witter, 1993). It has also been noted by Levy (1990) that, because the entorhinal cortex is so much larger than the hippocampus, the signal passing from the entorhinal cortex must be made as efficient as possible by reducing redundancy; this is consistent with the clustering or compression function we propose.

In the next section, we present a connectionist model which incorporates this entorhinal network together with

a long-term memory network. We propose that this model corresponds to the functioning that remains after an H lesion, which selectively damages hippocampal formation without otherwise disrupting the entorhinal cortex. The behavior of this model compares to the behavior of H-lesioned animals on several classical conditioning tasks for which stimulus–stimulus redundancy compression is expected to be critical.

A MODEL OF ENTORHINAL-CORTICAL FUNCTION

The intact model of Figure 1A contains both a network representing the hippocampal region, and a network representing long-term memory processes in the cerebral or cerebellar cortex. The hippocampal-region function in this top-down model is undifferentiated, and the H^{++}-lesioned model of Figure 1B, which deletes the hippocampal network, therefore compares to a broad lesion which disables all hippocampal-region processing. We now introduce a third model, which is intermediate between these two and represents the processing presumed to remain after selective hippocampal formation damage which does not include the entorhinal cortex. This H-lesioned model is shown in Figure 6. In this model, the representations formed in the entorhinal network are acquired by the long-term memory network, just as in the intact model; now, however, these representations are biased only by the constraint of stimulus–stimulus redundancy compression.

Most early hippocampal-lesion studies involved aspiration of the hippocampus; in these lesions, the overlying entorhinal and other cortices were generally damaged. Even more selective lesion, such as electrolytic destruction of the hippocampus, which does not damage entorhinal cell bodies directly, still severs entorhinal fibers of passage through the hippocampus (cf. Honey & Good, 1993). More recently, new techniques such as ibotenic acid injection (e.g., Jarrard, 1989) have made it possible to destroy hippocampal cell bodies while leaving fibers of passage intact. These lesions include the dentate gyrus (cf. Davidson, McKernan, & Jarrard, 1993; Honey & Good, 1993)— but not the overlying parahippocampal structures, including the entorhinal cortex, perirhinal cortex, and parahippocampal cortex. The fact that ibotenate lesion often results in less severe impairment than the broader aspiration lesions indicates that at least some parahippocampal processing can survive selective hippocampal lesion (Eichenbaum et al., 1994; Jarrard, 1993; Zola-Morgan & Squire, 1993). It should be noted, however, that the converse is not true: selective entorhinal lesion effectively isolates the hippocampus by destroying its primary source of sensory input (Amaral & Witter, 1989) and the primary route for hippocampal output back to cortex (Swanson, 1979). Studies of parahippocampal region processing typically involve lesion to several structures within the area, making it difficult to specify the precise function of the en-
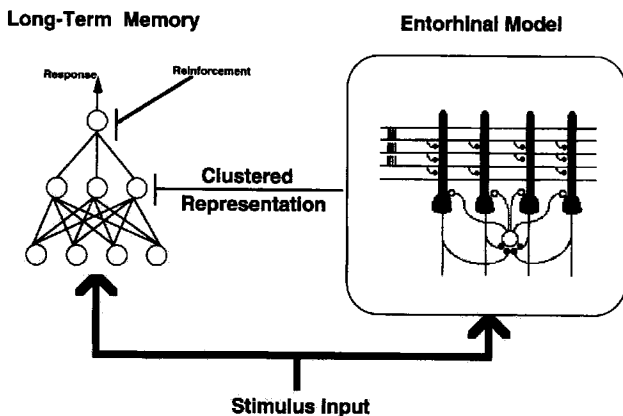
Long-Term Memory   **Entorhinal Model**

Stimulus Input

Figure 6. The H-lesioned model: The entorhinal network, on the right, performs unsupervised stimulus–stimulus redundancy compression or clustering. This is a subset of the representational functionality assumed in the hippocampal-region network of Figure 1A. The entorhinal representations are provided to the long-term memory network. This system is compared to behavior in animals with a selective hippocampal lesion which spares entorhinal processing (the H lesion).

torhinal cortex as distinct from other parahippocampal region structures. Apparently, the perirhinal cortex processes largely visual information, while the parahippocampal cortex processes spatial information; together, these two areas provide two-thirds of the input to the entorhinal cortex (Zola-Morgan & Squire, 1993). However, some processing in the parahippocampal and perirhinal cortices apparently can survive entorhinal damage (Murray, Gaffan, & Mishkin, 1993; Murray & Mishkin, 1986); therefore, although we concentrate here on the particular function of the entorhinal cortex, additional functionality is probably provided by other parahippocampal region structures.

Another important difference between ibotenate and aspiration lesion is that the former spares subiculum; although this structure, together with the dentate gyrus, is often considered to form a functional unit with the hippocampus proper (cf. Eichenbaum et al., 1994), its precise functions are undetermined. Therefore, the lesion which would be most precisely comparable to the selectively lesioned, entorhinal-cortical model would be an ibotenate lesion of the hippocampus together with dentate gyrus and subiculum, but sparing the entorhinal cortex. In the absence of empirical data regarding the effects of such a lesion, we will concentrate on comparing the entorhinal–cortical model to animals with ibotenate lesions of the hippocampus alone. Such animals presumably have residual entorhinal function like that assumed in the model; they may have additional function in their subiculum as well, which the model does not consider. For this reason, the following analysis focuses on several behaviors assumed to depend heavily on stimulus–stimulus

redundancy compression, for which the entorhinal cortex is assumed to be both necessary and sufficient.

It should be noted at the outset that the behavioral results shown for the H-lesioned model should not necessarily be quantitatively compared with the data shown in Figure 2, representing performance of the intact and H^{++}-lesioned models. Gluck and Myers's (1993) intact model is a top-down, abstract model, with learning procedures based on connectionist learning rules more closely related to traditional psychological theories of learning than to physiologically based rules for synaptic plasticity. Thus, while the intact model's behavior may be qualitatively comparable to that of normal animals, neither its learning rules nor its capacity are related to biological substrates. In contrast, the connectivity and learning rules found in the newer H-lesioned model presented here seek a closer rapprochement with biology. This implies that this H-lesioned model cannot necessarily be equated at a quantitative level of detail with the intact model in terms of capacity, functionality or speed. For example, while it is possible to say that both the intact and H-lesioned models can learn a simple conditioned discrimination, it is not appropriate to directly compare the models' relative speed to learn this discrimination. Rather, the nature of the model comparisons seen in Figures 7 through 10 will be based on within-group ordinal relationships between the difficulty of various training tasks compared with subsequent transfer tasks. As we shall see in the following sections, the H-lesioned model shows the same pattern of training and transfer effects as the intact model on tasks that depend uniquely on stimulus–stimulus redundancy compression. In contrast, for
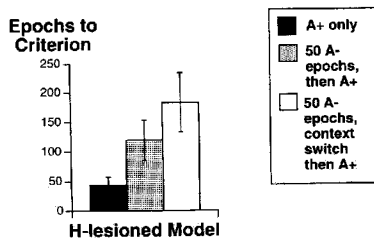
Epochs to Criterion

Figure 7. Latent inhibition in the H-lesioned model. Learning to respond to a cue, A, is retarded (reflected in more trials required to reach criterion responding) after 50 epochs of unreinforced preexposure to A than in a control condition with no preexposure. During the preexposure phase, stimulus–stimulus compression is assumed to increase generalization between A and the context, hindering later learning to respond to A but not to the context alone. Similarly, rats with selective (ibotenate) hippocampal lesions show a latent inhibition effect (Honey & Good, 1993). Latent inhibition is not abolished by a context shift between preexposure and acquisition; in fact, there is a slight (nonsignificant) increase in latent inhibition after context shift in the H-lesioned model. Similarly, selective ibotenate hippocampal lesions abolish the context sensitivity of latent inhibition in rats (Honey & Good, 1993).
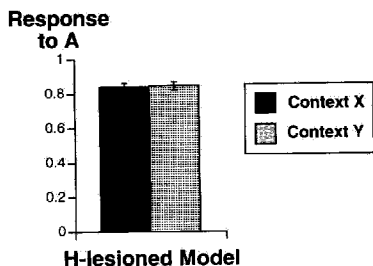
Response to A

Figure 8. Absence of response decrement with context shift in the H-lesioned model. After 200 epochs of training to respond to cue A in context X, there is little or no decrement in the response when A is presented in a novel context, Y. With no differentiation mechanisms, there is a net tendency to compress the representation of A with the context X in which it always appears. The task of responding selectively to A can be learned only by ignoring all but those aspects of the representation that are selectively active when A is present; these aspects ignore contextual information, and so will not be affected by a context shift. Similarly, selective (ibotenate) hippocampal lesions abolish the response decrement after context shift in rats (Honey & Good, 1993).

tasks that reflect the other representational biases shown in Table 1, the H-lesioned model behaves more like the H^{++}-lesioned model of Gluck and Myers (1993).

Ultimately we would like to develop an integrated and physiologically based intact model which would include dissociable entorhinal and hippocampal components. This would allow us to make quantitative comparisons among the models by showing that an initial intact model can be altered in different ways to produce behavioral consequences analogous to different lesion types. This goal,

however, is beyond the scope of the current paper. Here, in this paper, we have a more modest aim: we wish to show that the physiologically based model of the olfactory cortex developed by Granger and his colleagues provides a framework for interpreting Gluck and Myers's

(A)

Response to B

(B)

Diff(A+, B-)
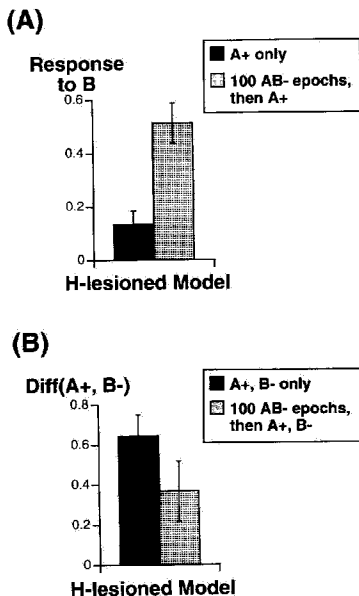
Figure 9. Behavioral results depending on stimulus–stimulus redundancy compression are maintained in the H-lesioned model, including sensory preconditioning (A) and compound preconditioning (B). These effects have not been tested in H-lesioned animals, and so represent novel predictions of the model.

Diff(H+, H-)
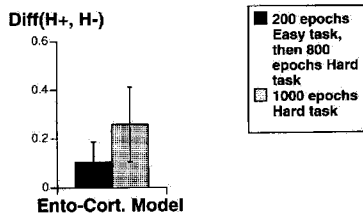
Figure 10. Easy–hard transfer, which requires other representational processes beyond stimulus–stimulus compression, is not maintained in the H-lesioned model. There is even a slight (nonsignificant) trend for learning on the hard task to be facilitated after pretraining on the hard—not easy—task, as indicated by a greater ability to respond differently to the stimuli in the hard task: Diff(H+, H−) is larger in the hard–hard condition than in the easy–hard condition.

(1993) redundancy compression as being a result of entorhinal and/or parahippocampal processing. Even with the limited hybrid model shown in Figure 6, it is possible, however, to make some important comparisons to behavioral analyses among different lesion groups, as described below.

Latent Inhibition

Figure 2E illustrates how the intact, but not H[+1]-lesioned, model shows a latent inhibition effect, whereby unreinforced preexposure to a cue, A, retards later acquisition of a response to A. As shown in Table 4, latent inhibition in the intact model results from the interplay of several representational processes during the preexposure phase; one is a bias for stimulus-outcome compression of cue A and the context, X, which both predict no particular salient outcome. There are also opportunities for stimulus–stimulus compression, because A always appears in X, as well as for stimulus–stimulus differentiation, since X does not always appear with A.

In contrast, the H-lesioned model has the capacity for neither stimulus-outcome differentiation nor stimulus–stimulus differentiation. As a result, only stimulus–stimulus compression of A and X occurs during the preexposure phase. The result is the same in both models—compression of the representations of A and X, as shown in Table 4—although additional mechanisms operate in the intact model. Since the representations of A and X are compressed in the H-lesioned model, there is still a latent inhibition effect: subsequent learning to discriminate A from X alone is retarded. During the subsequent acquisition phase, the intact model is assumed to redifferentiate the representations of A and X. The H-lesioned model cannot do this. The only way for it to solve the task is for the long-term memory network, as it acquires the compressed entorhinal representations, to learn selective responses based on the representations as they are. This can be done by weighting connections to the output nodes from just those internal layer nodes that are active (or not active) exactly when A is present. As long as at least some such nodes exist in the compressed representation, the task can still be learned. However, without the facilitating influence of differentiating between A and X during acquisition, this learning in the lesioned model is slower than in a control condition in which representations have not been previously compressed. As a result, the H-lesioned system shows a latent inhibition effect (Figure 7). Similarly, rats with selective, ibotenate hippocampal lesions show latent inhibition (Honey & Good, 1993).[1] In a taste-aversion paradigm, rats with ibotenate lesions even show increased latent inhibition, consistent with an interpretation that selective hippocampal lesion results in overcompression of cue and context (Reilly, Harley, & Revusky, 1993).[2]

Release From Latent Inhibition with Context Shift

The above analysis suggests that even though the H-lesioned and intact animals show similar compression of cue A and context X during unreinforced preexposure, the processes by which they subsequently learn to discriminate A and X are quite different. The intact model can use hippocampal-dependent stimulus-outcome differentiation, whereas the H-lesioned model can only search for features of the existing representation which already differentiate A and X alone. The difference in these processes is made evident during a transfer task, such as a shift to a new context, Y, between the preexposure and acquisition phases. Figure 2F shows that a context shift eliminates latent inhibition in the intact model.

By contrast, a context shift does not eliminate latent inhibition in the H-lesioned model, as shown in Figure 7; there is even a slight (nonsignificant) increase in the epochs needed to learn in the new context. During the preexposure phase, representational compression is assumed to result in increased generalization between cue A and context X, as well as simply reducing the representational resources allocated to encode for the presence of A in the environment. Shifting to a new context, Y, may eliminate the first problem but not the second. Learning to discriminate the presence of A, and respond selectively to it, still requires finding aspects of the representation that respond strongly when A is present, but not to the context alone when A is absent (or vice versa). This will still slow learning relative to a control condition with no preexposure. Consistent with this interpretation, a selective ibotenate hippocampal lesion suffices to eliminate the context sensitivity of latent inhibition in rats (Honey & Good, 1993).

Response Decrement After Context Shift

A similar distinction between representational mechanisms in the intact and H-lesioned models occurs during simple acquisition of a response to a cue A in context X. The representational processes that occur during A+ acquisition are schematized in Table 5. In the intact model, there is stimulus–stimulus compression (since A predicts X) and stimulus–stimulus differentiation (since X does not always predict A). There is also strong stimulus-outcome differentiation, since A predicts reinforcement but X alone does not. Early in training, before stimulus-

Table 4
Representational Changes During Unreinforced Preexposure

	Bias to Compress	Bias to Differentiate
Stimulus–stimulus regularity	$A \Rightarrow X$	$X \Rightarrow A$, $X \Rightarrow$ not A
Stimulus-outcome regularity	$AX \Rightarrow$ no US, X alone $\Rightarrow$ no US	

Note.—Representational changes during preexposure to cue A in context X. Shaded areas are representational biases that are found in the intact, but not the H-lesioned, model. In both models, A appears only in context X, and so there is a bias to compress the representations of A and X. In the intact model, there are two additional processes. Since X often appears without A, stimulus–stimulus differentiation occurs; since A and X both predict no particular outcome, stimulus-outcome compression of A and X also occurs. In both the intact and H-lesioned models, the net effect is to compress the representations of A and X, increasing generalization between them and slowing subsequent learning to discriminate them. Thus, both intact and H-lesioned models show a latent inhibition effect.

Table 5
Representational Changes During Response Acquisition

	Bias to Compress	Bias to Differentiate
Stimulus–stimulus regularity	$A \Rightarrow X$	$X \Rightarrow A, X \Rightarrow \text{not } A$
Stimulus–outcome regularity		$XA \Rightarrow US,$ $X \text{ alone} \Rightarrow \text{no } US$

Note—Representational changes during acquisition of a response to cue A in context X. Shaded areas are representational biases that are found in the intact, but not the H-lesioned, model. In both models, A appears only in context X, some stimulus–stimulus compression of A and X occurs. In the intact model, there are two additional processes. Since X often appears without A, there is stimulus–stimulus differentiation of A and X; since A predicts reinforcement (the unconditioned stimulus, or US) but X alone does not, there is stimulus–outcome differentiation of A and X. Thus, there is a net tendency to *compress* the representations of A and X in the H-lesioned model, but an eventual net tendency to *differentiate* representations of A and X in the intact model.

outcome regularities are learned, enough compression may occur so that there is a response decrement if A is presented in a new context, Y, as shown in Figure 2D. As training progresses, and stimulus-outcome differentiation comes to dominate, the response may remain strong in a new context.

In the H-lesioned model, by contrast, only stimulus–stimulus compression occurs, as shown in Table 5. By its nature, this compression impedes the task of learning to differentiate cue and context and respond selectively when A is present. In order to learn this task, the long-term-memory network must map from these compressed representations to responses, focusing on those aspects of the representation which still discriminate between the presence and absence of A. The nodes in the internal layer that are strongly active only when A is present (or only when A is absent) will be weighted most heavily. By selectively weighting these nodes, the network in effect ignores all other aspects of the representation, including contextual information. As a result, the learned response is relatively insensitive to subsequent contextual shifts, as shown in Figure 8. This is consistent with the failure of context shift to disrupt learned responding in rats with selective ibotenate hippocampal lesion (Honey & Good, 1993).

This interpretation is also consistent with results indicating that dorsal hippocampal lesions, which spare the entorhinal cortex, affect contextual fear conditioning. In normal rats, forward pairing of a tone stimulus with footshock results in conditioned freezing to the experimental context. Lesioned rats have been shown to condition to the tone but not to the context, showing a selective impairment in contextual conditioning (Phillips & LeDoux, 1994). Like in the context-shift experiments, this indicates that H-lesioned animals tend to ignore contextual information if a predictive cue is present, just as our theory predicts.

Other Behavioral Predictions

In addition to contextual tasks, the selectively lesioned, entorhinal-cortical model can be applied to the full domain of tasks modeled by the original intact and H[++]-

lesioned models in Gluck & Myers (1993). To the extent that these tasks can be neatly classified as depending primarily on stimulus–stimulus redundancy compression or predictive differentiation mechanisms, they should be spared or altered in the entorhinal-cortical model.

For example, the intact model shows sensory preconditioning (Figure 2A) and compound preconditioning (Figure 2B), both explained in terms of stimulus–stimulus redundancy compression during preexposure to the compound AB. The H-lesioned model, which also has stimulus–stimulus compression, continues to show both effects (Figure 9). This contrasts with the failure of the H[++]-lesioned model to show either effect. Sensory preconditioning has been shown to be eliminated by fornix lesion, which does not explicitly damage the entorhinal cortex (Port & Patterson, 1984). However, fornix lesion disrupts a hippocampal input–output pathway, and the functional results may be quite different from those following direct hippocampal ablation (Jarrard, 1993, 1994). It is not yet clear whether sensory preconditioning will survive a selective (ibotenic) hippocampal lesion, as our hypothesis predicts. Compound preconditioning has not been explored after any form of hippocampal damage, and so the behavior of both the H[++]-lesioned and H-lesioned models represent novel predictions which remain to be tested empirically. The models expect that both effects should be eliminated by broad hippocampal-region damage which includes the entorhinal cortex, but be spared by selective hippocampal damage which does not include the entorhinal cortex.

Conversely, other effects in the intact model are assumed to depend on processes not available in the H-lesioned model. The easy–hard transfer of Figure 2C is not obtained in the H-lesioned model (Figure 10) because it depends on stimulus-outcome differentiation mechanisms. If anything, pretraining on the easy task results in slightly slower learning of the hard task than does pretraining on the hard task itself, although the difference between the two conditions is not statistically significant. Again, to our knowledge, this effect has not been explored in animals with any form of hippocampal lesion.

The predictions described in this section remain to be tested empirically, comparing selective (ibotenate) hippocampal lesion with broader damage that includes entorhinal cortex. While there have been many studies of selective hippocampal lesion (e.g., Eichenbaum et al., 1994; Jarrard, 1993; Zola-Morgan & Squire, 1993), most of these have not involved simple conditioning tasks—with the exceptions noted above. It should also be mentioned that there are other conditioned behaviors which are known to be disrupted by broad H[++] lesion but not by selective H lesion, including negative patterning and conditional discrimination (Jarrard, 1993). These behaviors are not as neatly attributable to compression and differentiation processes as are those described above, and therefore it is less clear what our model predicts.

Negative patterning, for example, involves learning to respond to two components (A+, B−) but not their compound (AB−). A solution to this task requires forming

compound representations (AB, etc.) which our hypothesis predicts would occur in the entorhinal cortex. However, in the example of negative patterning, the solution further requires that the compound representation be distinguished from the component representations; negative patterning requires not just compression but also a critical differentiation of components and compounds. This is beyond the scope of our simple model of compression; as A and B become compressed, the components, by definition, become less distinguishable. As a result, the model responds in the same way to A, B, and AB, and the H-lesioned model incorrectly shows severe difficulty in learning negative patterning (simulations not shown). Similar causes lead to an incorrect prediction of disruption on conditional discrimination (simulations not shown). Therefore, if, as we hypothesize, the entorhinal cortex performs redundancy compression, there must be additional processes present in the H-lesioned animal—but not in our H-lesioned model—which allow configuration to occur. The H-lesioned animal retains a number of hippocampal-region structures besides the entorhinal cortex, including the subiculum (and often parts of the hippocampus itself). It also retains perirhinal and parahippocampal cortices. Each of these structures doubtless contributes an important component of hippocampal-region processing and it will doubtless be an important challenge for future modeling to develop a physiologically based interpretation of their functional role. With the present model, however, it is important to stress again that our H-lesioned model represents the entorhinal processing which is assumed to survive selective hippocampal lesion, and does not address the possible roles of these other related structures which may survive the H lesion.

RELATIONSHIP TO OTHER THEORIES OF ENTORHINAL–HIPPOCAMPAL FUNCTION

Eichenbaum, Otto, and Cohen (1994) have also addressed the selective contribution of the entorhinal cortex to hippocampal-region processing. They have suggested that the parahippocampal region (including the entorhinal cortex) functions as an intermediate-term buffer (Eichenbaum et al., 1994), while the hippocampal formation mediates the representational processing which allows for the generalization of learned associations in novel tasks (Eichenbaum et al., 1994). Although intermediate-term memory differs from the stimulus–stimulus compression function we have proposed for the entorhinal cortex, the two functions are complementary. In real-world situations, events which conceptually co-occur may not actually coinitiate or coterminate; intermediate-term storage of recent events might allow the system to cluster stimuli which reliably occur close together in time. Conversely, the intermediate-term store probably has to perform some sort of clustering to reduce the information passing through it. Analyses of the anatomical structure reveal that the entorhinal cortex contains many more afferents than efferents, and this supports the suggestion that the entorhinal cortex compresses information into a more efficient signal (Levy, 1990). Thus, although our hypothesis and that of Eichenbaum et al. (1994) suggest different functions for the entorhinal cortex, we feel that these functions are complementary and even interdependent (Gluck, Myers, & Goebel, 1994).

The remaining relational processing which Eichenbaum et al. presume occurs in the hippocampus can similarly be related to our own hypothesis. They suggest that hippocampal-mediated representations emphasize the relations between stimuli, and that they can be applied flexibly in novel situations (Eichenbaum, Cohen, Otto, & Wible, 1992; Eichenbaum, Otto, & Cohen, 1992). This is consistent with our suggestion that the hippocampus develops new stimulus representations which emphasize predictive features (and implicitly de-emphasize irrelevant information). This may allow the flexible use of these representations in new contexts whose irrelevant features may differ from the initial learning environment.

More recently, Bunsey and Eichenbaum (1993) have suggested that the parahippocampal region also mediates the "fusion" of co-occurring or nearly coincident stimuli; this process is functionally identical to the redundancy compression function described here and in Gluck and Myers (1993). It is interesting to note that whereas their "fusion" theory derives from behavioral observations comparing paired-associate learning in intact, hippocampal-lesioned, and parahippocampal-lesioned animals, our similar "compression" theory arises from an integration of both physiologically based and behaviorally based computational models of hippocampal-region function. The convergence of these two widely different approaches to theory development is hopefully a sign that stimulus compression is a useful and accurate description of parahippocampal-region function.

We have not yet addressed the anatomical substrates which might underlie the remaining representational processing assumed to occur in the hippocampus—namely, altering altering stimulus representations based on stimulus-outcome regularities. Other workers, however, have considered bottom-up models of hippocampal function. Treves and Rolls (1994) have suggested that hippocampal field CA3 functions as an autoassociator which can reconstruct stored patterns from partial or noisy versions; O'Reilly and McClelland (1994) have assumed similar functionality in CA3. Such an autoassociator is related to the more powerful autoencoder which we have envisioned as a model of the aggregate hippocampal region (Gluck & Myers, 1993). McNaughton and Nadel have gone further to assume that the autoassociator in the hippocampus can additionally predict the next state in a transition matrix (McNaughton & Nadel, 1990); a simple example of this prediction occurs in our autoencoding hippocampal-region network, which learns to output a prediction of the future in terms of an anticipatory response (the CR).

Along similar lines, Lynch and Granger (1992) have suggested that specific physiological characteristics of plasticity in hippocampal field CA1 may yield the func-

tion of anticipatory signaling of expected continuations of sequences. This idea that the hippocampus functions as a predictor of future events has also been explored by Gray (1985), Levy (1985), Treves and Rolls (1992), and others. If, as we have suggested, the hippocampus constructs new representations based on the predictive value of stimuli, it would be reasonable to consider that it might further be able to reconstruct a representation of the actual predicted future event and compare that with what was actually experienced. In sum, therefore, a number of bottom-up models of the hippocampus suggest it could be quite capable of performing the stimulus-outcome differentiation which our model assumes and which would not be mediated by the entorhinal cortex.

Other researchers have examined the hypothesis that the hippocampus represents a fast "one-trial" learning system distinct from the slow, incremental learning that occurs in the cortex (McClelland et al., 1994). In light of this, it is interesting to note that learning in our proposed entorhinal model is also very fast, with at least coarse clusters developing after as little as a single exposure to a stimulus (see also Granger et al., 1989). Such a system might be relevant during the kinds of recognition tasks that are impaired in humans after hippocampal-region damage (e.g., Haist, Musen, & Squire, 1991). However, consolidation of memories from the hippocampus to the cortex is usually meant to imply declarative (or episodic) learning; this is fundamentally different from the incremental conditioned learning which our model addresses. Although several researchers have begun to develop computational models that address the issues of fast versus slow learning and the role of the hippocampus in the consolidation of long-term memories (Alvarez & Squire, 1994; McClelland et al., 1994; Murre, 1994), it remains an important challenge for the future to relate the phenomena and theory of memory consolidation in episodic learning to the range of representational deficits observed in conditioning studies of hippocampal-lesioned animals.

In other modeling, Schmajuk and his colleagues (e.g., Schmajuk & DiCarlo, 1992) have suggested that the special circuitry of the hippocampal region allows it to compute error backpropagation, a powerful form of learning algorithm which we have also used in our intact model. In particular, they argue that the hippocampal region (1) forms and stores configural associations, and (2) computes and broadcasts an aggregate error signal which is necessary for stimulus competition of the kind described by Rescorla and Wagner's (1972) model of associative learning. Their models account for an extensive range of behavioral data from the classical conditioning domain, including many of the same phenomena addressed by our model.

The Schmajuk and DiCarlo (1992) model, however, differs from ours in some specific and important predictions. First, while we have used backpropagation as a convenient approximation to derive the representational recoding suggested by the underlying qualitative theory, Schmajuk and DiCarlo interpret the backpropagation algorithm more literally and employ it as an integral part of their model. Thus, the Schmajuk–DiCarlo model is more concerned than ours with showing that a biologically plausible implementation of backpropagation could exist in the hippocampus. A second major difference between the models is that the Schmajuk–DiCarlo model predicts that the hippocampus is necessary for both the acquisition and the storage of configural associations. Our model does not make such task-specific predictions. Instead, our model argues that hippocampal lesion disrupts representational processing; configural associations are simply an example of a task which generally requires new stimulus representations and therefore is often disrupted by hippocampal lesion. However, our H^{++}-lesioned model does have (fixed) stimulus representations, and there is always some small probability that these random representations suffice to solve any given particular task. By this argument, there may be instances in which configural learning is not disrupted or instances in which configural tasks are learnable by the hippocampal-lesioned animal. Although this is an important conceptual distinction between the models, it may be difficult to test, because it requires a priori knowledge of how stimuli are represented by an individual animal. Furthermore, although our model predicts that configural associations are eventually stored outside the hippocampal region, and therefore become hippocampal-independent, the consolidation period required for this transfer is likewise generally indeterminate.

Although the Schmajuk–DiCarlo model can account for many conditioning data, it fails to address the hippocampal-dependent phenomena that involve such stimulus–stimulus associations as latent inhibition and sensory preconditioning, phenomena which are addressed by the Gluck–Myers model and central to the analyses we have presented here and elsewhere (Gluck & Myers, 1993; Myers & Gluck, 1994).

The Schmajuk–DiCarlo model also predicts that the hippocampus is required for stimulus competition, and therefore expects such effects as blocking, conditioned inhibition, and overshadowing to be hippocampal-dependent. Contrary to the expectation of the Schmajuk–DiCarlo model, conditioned inhibition is hippocampal-independent (Solomon, 1977), although overshadowing has been shown to depend on an intact hippocampus (Schmajuk, Spear, & Isaacson, 1983). In the Gluck and Myers (1993) model, stimulus competition effects occur both in the hippocampal network, as a result of representational changes, and in the long-term memory network, as a result of the reinforcement modulation underlying the error-correcting learning procedure. The H^{++}-lesioned model maintains the latter kind of stimulus competition, predicting that conditioned effects which can be explained in terms of either kind of competition may be weakened by hippocampal lesion, but will not necessarily be eliminated. Two example effects are blocking (Kamin, 1969) and overshadowing (Kehoe & Gormezano, 1980), in which strongly associated cues prevent other cues from gaining associative strength. Both effects are shown in the fully lesioned model, predicting that both effects should not

be eliminated by broad hippocampal-region damage (Gluck & Myers, 1993). These predictions contrast with the assumption in the Schmajuk-DiCarlo model that the hippocampus is critical for all forms of stimulus selection, and therefore that hippocampal damage should eliminate both blocking and overshadowing.

Contradictory empirical results have been reported for these phenomena. For example, early data suggested that blocking depends on an intact hippocampus (Solomon, 1977), while more recent studies have had difficulty replicating this result (Garrud et al., 1984). The assumption that the hippocampus is critical for blocking is also at odds with data and theory which argue for a sufficient cerebellar circuit for blocking and overshadowing in motor-reflex conditioning (see Donegan, Gluck, & Thompson, 1989; Gluck, Goren, et al., 1994). In contrast to the implications of the Schmajuk-DiCarlo model, the aforementioned anatomical and behavioral data seem to support the idea, made explicit in our corticohippocampal theory, that the hippocampus is one, but not the only, substrate of stimulus selection in associative learning.

More recently, Schmajuk (1994; Schmajuk & Blair, 1993) has suggested the particular contribution of the entorhinal cortex to Schmajuk and DiCarlo's (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 the entorhinal cortex, should eliminate the configural but not the stimulus-competition function. Such empirical data as exist 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 Schmajuk and Blair proposed 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 and 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 which address the selective role of the entorhinal cortex in stimulus competition and in stimulus–stimulus clustering are indicated to properly evaluate these two models.

SUMMARY AND DISCUSSION

Gluck and Myers (1993) proposed that the hippocampal region is critically involved in the construction of internal representations of stimuli. These representations are constrained to compress or differentiate information based on both stimulus–stimulus and stimulus-outcome regularities. In turn, representational similarity affects the amount to which associations generalize across stimuli. This simple representational account of hippocampal-region func-

tion is sufficient to accurately predict a wide range of conditioned behaviors in the intact and lesioned animal.

The present paper presents the hypothesis that the proposed representational function can be subdivided and distributed among different anatomical structures within the hippocampal region. Specifically, anatomical and physiological clues suggest that stimulus–stimulus redundancy compression occurs in the entorhinal cortex; the remaining representational changes would then occur in other hippocampal-region structures. We have implemented this hypothesis by considering an unsupervised competitive network representing superficial entorhinal cortex clustering mechanisms; the new clustered representations formed by this network are then made available to a long-term memory network which can acquire these representations but not construct its own. This model can be compared to behavior in an animal with a selective hippocampal lesion which does not damage the entorhinal cortex (the H lesion), and predicts the survival of such behaviors as latent inhibition and sensory preconditioning, which can arise solely from stimulus–stimulus redundancy compression but do not require other hippocampal-mediated representational changes. As described earlier, empirical data show that such a limited lesion does not destroy latent inhibition, but does disrupt contextual processing effects, consistent with the model predictions. However, other of the model's behavioral predictions remain to be tested. For example, H-lesioned animals are expected to continue to show sensory preconditioning and compound preconditioning, but not easy–hard transfer.

A cornerstone of our hypothesis is that entorhinal cortex and sensory cortices may be performing similar stimulus–stimulus redundancy compression functions. Whereas sensory cortices deal with a single modality, the entorhinal cortex would be well placed to compress stimuli across modalities or compress cross-modal features of a single stimulus. This implies that entorhinal damage may have dramatically different effects, depending on whether a task uses unimodal or multimodal stimuli. For example, sensory preconditioning is assumed to depend on redundancy compression of two cues during a phase of unreinforced preexposure. If the two cues are from different modalities, this effect should be sensitive to entorhinal damage; if the two cues are from the same modality, compression processes within the sensory cortex might well be able to produce the effect. This aspect of task demands has not received a great deal of empirical attention in lesioned animals, although it has been well studied in behavioral studies in intact animals. An interesting set of empirical studies would be to test sensory preconditioning or compound preconditioning after selective (ibotenate) hippocampal lesion, comparing unimodal and polymodal task versions.

An obvious limitation of our current model is that although we have constructed a bottom-up model of the entorhinal cortex, allowing us to implement an H-lesioned model, we have not yet presented a bottom-up model of hippocampus proper. This is clearly a direction for future

work. One of the important issues here is whether it is possible for the system to operate in a truly serial fashion, with the hippocampus operating on previously compressed entorhinal data, or whether there must be reciprocal pathways allowing hippocampal processing to influence entorhinal compression. It seems likely that at least some reciprocity is necessary to ensure that necessary, predictive information is not overcompressed and lost by unsupervised entorhinal processing. A second important issue will be the role of subcortical input on hippocampal processing. The hippocampus receives an important subcortical input via the fornix, and also outputs through the fornix. One purpose of this exchange may be to modulate hippocampal processing between learning or storage and retrieval states (cf. Haas, 1983; Hasselmo & Schnell, 1994). One advantage of developing a bottom-up model of the hippocampus that addresses fornical influence will be to generate predictions regarding the effects of fornix lesion. At this point, it seems likely that fornix transection results not in eliminating hippocampal activity but in disrupting it, and that this disruption may be more severe than outright lesion (Berry & Thompson, 1979; Jarrard, Okaichi, Steward, & Goldschmidt, 1984). This idea is broadly consistent with our model, in which simple tasks may be learnable without hippocampal-region involvement, but disrupted hippocampal representations could prevent stable solutions from being constructed altogether (cf. Gluck & Myers, 1993). A bottom-up model of the hippocampus that can be incorporated with our entorhinal model into a more complete model of intact processing will allow some of these issues to be explored.

Perhaps the primary implication of this paper is that further work is clearly needed to more carefully reexamine the effects of selective lesions of hippocampal formation versus lesions of surrounding cortical tissue on simple conditioning tasks. To date, there has been a dearth of such empirical studies. Selective-lesion studies have instead largely concentrated on such higher level tasks as delayed nonmatching in monkeys (Zola-Morgan et al., 1992), odor discrimination learning in rats (Eichenbaum et al., 1994), and conditional and spatial learning in rats (Jarrard, 1993). Animals display dramatically different behaviors on these tasks when the entorhinal cortex is lesioned and when the hippocampus alone is damaged. However, our model is currently too limited to address these higher level tasks. Our model does not address temporal information, and therefore cannot simulate delays between stimulus presentations. It also does not address the issues of hippocampal involvement in one-trial learning such as may underlie episodic or declarative memory (Squire, 1987) or the consolidation period during which episodic memories become independent of hippocampal mediation (McClelland et al., 1994; Squire, 1987). Extending the present theoretical framework to incorporate and address these other phenomena is an important challenge for future work.

The simplest kind of learning for which selective lesion studies do exist is forced-choice learning, a type of operant conditioning in which the subject is required to respond to a stimulus with one of a finite set of possible responses. Although this is more complex than the classical conditioning we have considered, it is still a form of incremental learning which should be possible with a few superficial extensions to our model. In particular, Eichenbaum and his colleagues have found that while rats with hippocampal damage (via fornix lesion) are able to learn normally if stimuli are presented successively, they are greatly impaired if stimuli are presented simultaneously (Eichenbaum, Fagan, Mathews, & Cohen, 1988). Eichenbaum has suggested that these results imply that the hippocampal region is critical for learning relationships between stimuli (Eichenbaum et al., 1988; Eichenbaum, Mathews, & Cohen, 1989). Since the difference between successive and simultaneous discrimination is one of stimulus–stimulus co-occurrence, we expect that stimulus–stimulus compression and differentiation in our model may have a critical influence on learning these tasks. Again, this remains as a challenge for further investigation and theoretical analysis.

Even within the limited domain of classical conditioning, we have attempted to demonstrate that top-down behaviorally constrained computational modeling and bottom-up physiologically motivated models can interact and inform one another. We take the convergent implications of these distinct methodologies as suggestive of how both approaches can contribute to the development of a unified understanding of hippocampal-region function in learning and memory.

REFERENCES

ALVAREZ, P., & SQUIRE, L. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences*, 91, 7041-7045.

AMARAL, D. G., ISHIZUKA, N., & CLAIBORNE, B. (1990). Neurons, numbers and the hippocampal network. In J. Storm-Mathisen, J. Zimmer, & O. P. Ottersen (Eds.), *Progress in brain research* (pp. 1-11). Amsterdam: Elsevier.

AMARAL, D. G., & WITTER, M. P. (1989). The three-dimensional organization of the hippocampal formation: A review of anatomical data. *Neuroscience*, 31, 571-591.

AMBROS-INGERSON, J., GRANGER, R., & LYNCH, G. (1990). Simulation of paleocortex performs hierarchical clustering. *Science*, 247, 1344-1348.

ANTELMAN, S., & BROWN, T. (1972). Hippocampal lesions and shuttlebox avoidance behavior: A fear hypothesis. *Physiology & Behavior*, 9, 15-20.

ANTON, P., GRANGER, R., & LYNCH, G. (1992). Temporal information processing in synapses, cells and circuits. In T. McKenna, J. Davis, & S. Zornetzer (Eds.), *Single neuron computation* (pp. 291-313). New York: Academic Press.

ANTON, P., GRANGER, R., & LYNCH, G. (1993). Simulated dendritic spines influence reciprocal synaptic strengths and lateral inhibition in the olfactory bulb. *Brain Research*, 628, 157-165.

ANTON, P., LYNCH, G., & GRANGER, R. (1991). Computation of frequency-to-spatial transform by olfactory bulb glomeruli. *Biological Cybernetics*, 65, 407-414.

BERRY, S., & THOMPSON, R. (1979). Medial septal lesions retard classical conditioning of the nictitating membrane response in rabbits. *Science*, 205, 209-211.

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. *Journal of Physiology*, 232, 331-356.

BOUTON, M., & BROOKS, D. (1993). Time and context effects on performance in a Pavlovian discrimination reversal. *Journal of Experimental Psychology: Animal Behavior Processes*, **19**, 165-179.

BUNSEY, M., & EICHENBAUM, H. (1993). Critical role of the parahippocampal region for paired-associate learning in rats. *Behavioral Neuroscience*, **107**, 740-747.

BUZSAKI, G. (1989). Two-stage model of memory-trace formation: A role for "noisy" brain states. *Neuroscience*, **31**, 551-570.

COULTRIP, R., GRANGER, R., & LYNCH, G. (1992). A cortical model of winner-take-all competition via lateral inhibition. *Neural Networks*, **5**, 47-54.

DAVIDSON, T., McKERNAN, M., & JARRARD, L. (1993). Hippocampal lesions do not impair negative patterning: A challenge to configural association theory. *Behavioral Neuroscience*, **107**, 227-234.

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 Research*, **600**, 327-330.

DONEGAN, N. H., GLUCK, M. A., & THOMPSON, R. F. (1989). Integrating behavioral and biological models of conditioning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 23, pp. 109-156). New York: Academic Press.

EICHENBAUM, H. (in press). The hippocampal system and declarative memory in humans and animals: Experimental analysis and historical origins. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994*. Cambridge MA: MIT Press.

EICHENBAUM, H., COHEN, N. J., OTTO, T., & WIBLE, C. (1992). Memory representation in the hippocampus: Functional domain and functional organization. In L. R. Squire, G. Lynch, N. M. Weinberger, & J. L. McGaugh (Eds.), *Memory organization and locus of change* (pp. 163-204). Oxford: Oxford University Press.

EICHENBAUM, H., FAGAN, A., MATHEWS, P., & COHEN, N. J. (1988). Hippocampal system dysfunction and odor discrimination learning in rats: Impairment or facilitation depending on representational demands. *Behavioral Neuroscience*, **102**, 331-339.

EICHENBAUM, H., MATHEWS, P., & COHEN, N. J. (1989). Further studies of hippocampal representation during odor discrimination learning. *Behavioral Neuroscience*, **103**, 1207-1216.

EICHENBAUM, H., OTTO, T., & COHEN, N.J. (1992). The hippocampus—What does it do? *Behavioral & Neural Biology*, **57**, 2-36.

EICHENBAUM, H., OTTO, T., & COHEN, N. (1994). Two functional components of the hippocampal memory system. *Behavioral & Brain Sciences*, **17**, 449-518.

ESTES, W. (1958). Stimulus-response theory of drive. In M. Jones (Eds.), *Nebraska Symposium on Motivation* (pp. 35-69). Lincoln: University of Nebraska Press.

FREY, P., & SEARS, R. (1978). Models of conditioning incorporating the Rescorla-Wagner associative axiom, a dynamic attention process, and a catastrophe rule. *Psychological Review*, **85**, 321-340.

GARRUD, P., RAWLINS, J., MACKINTOSH, N., GOODALL, G., COTTON, M., & FELDON, J. (1984). Successful overshadowing and blocking in hippocampectomized rats. *Behavioral Brain Research*, **12**, 39-53.

GLUCK, M., & BOWER, G. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, **117**, 225-244.

GLUCK, M., GOREN, O., MYERS, C., & THOMPSON, R. (1994). *A higher-order recurrent network model of the cerebellar substrates of response timing in motor-reflex conditioning*. Manuscript submitted for publication.

GLUCK, M., & GRANGER, R. (1993). Computational models of the neural bases of learning and memory. *Annual Review of Neuroscience*, **16**, 667-706.

GLUCK, M., & MYERS, C. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus*, **3**, 491-516.

GLUCK, M., & MYERS, C. (in press). Representation and association in memory: A neurocomputational view of hippocampal function. *Current Directions in Psychological Science*.

GLUCK, M., MYERS, C., & GOEBEL, J. (1994). A computational perspective on dissociating hippocampal and entorhinal function (Response to Eichenbaum, et al.). *Behavioral & Brain Sciences*, **17**, 478-479.

GOOD, M., & HONEY, R. C. (1991). Conditioning and contextual retrieval in hippocampal rats. *Behavioral Neuroscience*, **105**, 499-509.

GORMEZANO, I., KEHOE, E. J., & MARSHALL, B. S. (1983). Twenty years of classical conditioning research with the rabbit. *Progress in Psychobiology & Physiological Psychology*, **10**, 197-275.

GRANGER, R., AMBROS-INGERSON, J., & LYNCH, G. (1989). Derivation of encoding characteristics of layer II cerebral cortex. *Journal of Cognitive Neuroscience*, **1**, 61-87.

GRANGER, R., & LYNCH, G. (1991). Higher olfactory processes: Perceptual learning and memory. *Current Biology*, **1**, 209-214.

GRAY, J. A. (1985). Memory buffer and comparator can share the same circuitry. *Behavioral & Brain Sciences*, **8**, 501.

GROSSBERG, S. (1976). Adaptive pattern classification and recoding: Part I. *Biological Cybernetics*, **23**, 121-134.

HAAS, H. L. (1983). Amine neurotransmitter actions in the hippocampus. In W. Seifert (Eds.), *Neurobiology of the hippocampus* (pp. 139-155). London: Academic Press.

HAIST, F., MUSEN, G., & SQUIRE, L. R. (1991). Intact priming of words and nonwords in amnesia. *Psychobiology*, **19**, 275-285.

HALL, G., & HONEY, R. (1990). Context-specific conditioning in the conditioned-emotional-response procedure. *Journal of Experimental Psychology: Animal Behavior Processes*, **16**, 271-278.

HASSELMO, M., & SCHNELL, E. (1994). Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1: Computational modeling and brain slice physiology. *Journal of Neuroscience*, **14**, 3898-3914.

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. *Behavioral Biology*, **12**, 421-444.

HONEY, R., & GOOD, M. (1993). Selective hippocampal lesions abolish the contextual specificity of latent inhibition and conditioning. *Behavioral Neuroscience*, **107**, 23-33.

HONEY, R., WILLIS, A., & HALL, G. (1990). Context specificity in pigeon autoshaping. *Learning & Motivation*, **21**, 125-136.

HULL, C. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.

JARRARD, L. E. (1989). On the use of ibotenic acid to lesion selectively different components of the hippocampal formation. *Journal of Neuroscience Methods*, **29**, 251-259.

JARRARD, L. E. (1993). On the role of the hippocampus in learning and memory in the rat. *Behavioral & Neural Biology*, **60**, 9-26.

JARRARD, L. E. (1994). A call for greater concern regarding the underlying anatomy: Response to Eichenbaum, et al., target article. *Behavioral & Brain Sciences*, **17**, 485-486.

JARRARD, L. E., & DAVIDSON, T. L. (1991). On the hippocampus and learned conditional responding: Effects of aspiration versus ibotenate lesions. *Hippocampus*, **1**, 107-117.

JARRARD, L. E., OKAICHI, H., STEWARD, O., & GOLDSCHMIDT, R. B. (1984). On the role of hippocampal connections in the performance of place and cue tasks: Comparisons with damage to hippocampus. *Behavioral Neuroscience*, **98**, 946-954.

JUNG, M., LARSON, J., & LYNCH, G. (1990). Role of NMDA and non-NMDA receptors in synaptic transmission in rat piriform cortex. *Experimental Brain Research*, **82**, 451-455.

KAMIN, L. (1969). Predictability, surprise, attention and conditioning. In B. Campbell & R. Church (Eds.), *Punishment and aversive behavior* (pp. 279-296). New York: Appleton-Century-Crofts.

KANTER, E., & HABERLY, L. (1990). NMDA-dependent inhibition of long-term potentiation in afferent and association fiber systems of piriform cortex in vitro. *Brain Research*, **525**, 175-179.

KAYE, H., & PEARCE, J. (1987). Hippocampal lesions attenuate latent inhibition and the decline of the orienting response in rats. *Quarterly Journal of Experimental Psychology*, **39B**, 107-125.

KEHOE, E., & GORMEZANO, I. (1980). Configuration and combination laws in conditioning with compound stimuli. *Psychological Bulletin*, **87**, 351-378.

KENDLER, H., KENDLER, T., & SANDERS, J. (1967). Reversal and partial reversal shifts with verbal material. *Journal of Verbal Learning & Verbal Behavior*, **6**, 117-127.

KIM, J. J., & FANSELOW, M. S. (1992). Modality-specific retrograde amnesia of fear. *Science*, **256**, 675-677.

KOHLER, C. (1986). Intrinsic connections of the retrohippocampal region in the rat brain. II. The medial entorhinal area. *Journal of Comparative Neurology*, **246**, 149-169.

KOHLER, C., WU, J., & CHAN-PALAY, V. (1985). Neurons and terminals in the retrohippocampal region in the rat brain identified by anti-GABA and anti-GAD immunocytochemistry. *Anatomical Embryology*, **173**, 35-44.

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

LAWRENCE, D. H. (1952). The transfer of a discrimination along a continuum. *Journal of Comparative & Physiological Psychology*, **45**, 511-516.

LAWRENCE, D. H., & MASON, W. A. (1955). Systematic behavior during discrimination reversal and change of dimensions. *Journal of Comparative & Physiological Psychology*, **48**, 1-7.

LEVY, W. (1985). An information/computation theory of hippocampal function. *Society for Neuroscience Abstracts*, **11**, 493.

LEVY, W. (1990). Hippocampal theories and the information/computation perspective. In L. Erinoff (Ed.), *Neurobiology of drug abuse: Learning and memory* (pp. 116-125). Rockville, MD: U.S. Department of Health and Human Services, National Institute on Drug Abuse.

LEVY, W. B., BRASSEL, S. E., & MOORE, S. D. (1983). Partial quantification of the associative synaptic learning rule of the dentate gyrus. *Neuroscience*, **8**, 799-808.

LUBOW, R. (1973). Latent inhibition. *Psychological Bulletin*, **79**, 398-407.

LUBOW, R. E., RIFKIN, B., & ALEK, M. (1976). The context effect: The relationship between stimulus preexposure and environmental preexposure determines subsequent learning. *Journal of Experimental Psychology: Animal Behavior Processes*, **2**, 38-47.

LYNCH, G. (1986). *Synapses, circuits and the beginnings of memory*. London: MIT Press.

LYNCH, G., & GRANGER, R. (1992). Variations in synaptic plasticity and types of memory in corticohippocampal networks. *Journal of Cognitive Neuroscience*, **4**, 189-199.

MACKINTOSH, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, **82**, 276-298.

MACKINTOSH, N. J., & LITTLE, L. (1970). An analysis of transfer along a continuum. *Canadian Journal of Psychology*, **24**, 362-369.

MACRIDES, F. (1975). Temporal relationships between hippocampal slow waves and exploratory sniffing in hamsters. *Behavioral Biology*, **14**, 295-308.

MACRIDES, F., EICHENBAUM, H., & FORBES, W. (1982). Temporal relationship between sniffing and the limbic (theta) rhythm during odor discrimination reversal learning. *Journal of Neuroscience*, **2**, 1705-1717.

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*. (Tech. Rep. No. PDP.CNS.94.1). Pittsburgh: Carnegie Mellon University.

McNAUGHTON, B., & NADEL, L. (1990). Hebb-Marr networks and the neurobiological representation of action in space. In M. Gluck & D. Rumelhart (Eds.), *Neuroscience and connectionist theory* (pp. 1-63). Hillsdale, NJ: Erlbaum.

MISHKIN, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society of London: Series B*, **298**, 85-92.

MONAGHAN, D., & COTMAN, C. (1985). Distribution of N-methyl-D-aspartate-sensitive L-[³H]glutamate-binding sites in rat brain. *Journal of Neuroscience*, **5**, 2909-2919.

MURRAY, E., GAFFAN, D., & MISHKIN, M. (1993). Neural substrates of visual stimulus-stimulus association in rhesus monkeys. *Journal of Neuroscience*, **13**, 4549-4561.

MURRAY, E., & MISHKIN, M. (1986). Visual recognition in monkeys following rhinal cortical ablations combined with either amygdalectomy or hippocampectomy. *Journal of Neuroscience*, **6**, 1991-2003.

MURRE, J. (1994). *A model for amnesia*. Unpublished manuscript. Cambridge, U.K: Medical Research Council—Applied Psychology Unit.

MYERS, C., & GLUCK, M. (1994). Context, conditioning and hippocampal re-representation. *Behavioral Neuroscience*, **108**, 835-847.

NADEL, L. (1992). Multiple memory systems: What and why. *Journal of Cognitive Neuroscience*, **4**, 179-188.

O'REILLY, R., & McCLELLAND, J. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a tradeoff. *Hippocampus*, **4**, 661-682.

PARKER, D. (1985). *Learning logic* (Report). Cambridge, MA: MIT, Center for Computational Research in Economics and Management Science.

PAVLOV, I. (1927). *Conditioned reflexes*. London: Oxford University Press.

PEARCE, J., & HALL, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, **87**, 532-552.

PENICK, S., & SOLOMON, R. (1991). Hippocampus, context and conditioning. *Behavioral Neuroscience*, **105**, 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., MIKHAIL, A., & PATTERSON, M. (1985). Differential effects of hippocampectomy on classically conditioned rabbit nictitating membrane response related to interstimulus interval. *Behavioral Neuroscience*, **99**, 200-208.

PORT, R., & PATTERSON, M. (1984). Fimbrial lesions and sensory preconditioning. *Behavioral Neuroscience*, **98**, 584-589.

PRICE, J. (1973). An autoradiographic study of complementary laminar patterns of termination of afferent fiber to the olfactory cortex. *Journal of Comparative Neurology*, **150**, 87-108.

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. *Behavioral Neuroscience*, **107**, 996-1004.

RESCORLA, R., & WAGNER, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

RILEY, D. (1968). *Discrimination learning*. Boston: Allyn & Bacon.

ROMAN, F., STAUBLI, U., & LYNCH, G. (1987). Evidence for synaptic potentiation in a cortical network during learning. *Brain Research*, **418**, 221-226.

RUMELHART, D., HINTON, G., & WILLIAMS, R. (1986). Learning internal representations by error propagation. In D. Rumelhart & J. Mc-Clelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 318-362). Cambridge, MA: MIT Press.

RUMELHART, D., & ZIPSER, D. (1985). Feature discovery by competitive learning. *Cognitive Science*, **9**, 75-112.

SCHMAJUK, N. (1994). Stimulus configuration, classical conditioning, and spatial learning: Role of the hippocampus. In *World Congress on Neural Networks* (Vol. 2, pp. II723-II728). San Diego: INNS Press.

SCHMAJUK, N., & BLAIR, H. (1993). Stimulus configuration, spatial learning and hippocampal function. *Behavioral Brain Research*, **59**, 1-15.

SCHMAJUK, N., & DiCARLO, J. (1992). Stimulus configuration, classical conditioning and hippocampal function. *Psychological Review*, **99**, 268-305.

SCHMAJUK, N., & MOORE, J. (1988). The hippocampus and the classically conditioned nictitating membrane response: A real-time attentional model. *Psychobiology*, **16**, 20-35.

SCHMAJUK, N., SPEAR, N., & ISAACSON, R. (1983). Absence of overshadowing in rats with hippocampal lesions. *Physiological Psychology*, **11**, 59.

SCHOUSBOE, A., BACHEVALIER, J., BRAAK, H., HEINEMANN, U., NITSCH, R., SCHRODER, H., & WETMORE, C. (1993). Structural correlates and cellular mechanisms in entorhinal-hippocampal dysfunction. *Hippocampus*, **3**, 293-302.

SOLOMON, P. (1977). Role of the hippocampus in blocking and conditioned inhibition of the rabbit's nictitating membrane. *Journal of Comparative & Physiological Psychology*, **91**, 407-417.

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. *Journal of Comparative & Physiological Psychology*, **89**, 1192-1203.

SQUIRE, L. (1987). *Memory and brain*. New York: Oxford University Press.

SQUIRE, L. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, **99**, 195-231.

STANTON, P., & SEJNOWSKI, T. (1989). Storing covariance by the associative long-term potentiation and depression of synaptic strengths in the hippocampus. In D. Touretzky (Eds.), *Advances in neural information processing I* (pp. 394-401). San Mateo, CA: Morgan Kaufman.

SUTTON, R., & BARTO, A. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, **88**, 135-170.

SWANSON, L. (1979). The hippocampus—New anatomical insights. *Trends in Neurosciences*, **2**, 9-12.

TERRACE, H. (1963). Discrimination learning with and without "errors." *Journal of the Experimental Analysis of Behavior*, **6**, 1-27.

THOMPSON, R. (1972). Sensory preconditioning. In R. Thompson & J. Voss (Eds.), *Topics in learning and performance* (pp. 105-129). New York: Academic Press.

TRABASSO, T., & BOWER, G. H. (1964). *Attention in learning: Theory and research*. New York: Wiley.

TREVES, A., & ROLLS, E. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus*, **2**, 189-200.

VAN HOESEN, G. (1982). The parahippocampal gyrus: New observations regarding its cortical connections in the monkey. *Trends in Neurosciences*, **5**, 345-350.

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 Research*, **95**, 1-24.

VON DER MALSBURG, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, **14**, 85-100.

WAGNER, A. (1978). Expectancies and the priming of STM. In S. Hulse, H. Fowler, & W. Honig (Eds.), *Cognitive processes in animal behavior*. Hillsdale, NJ: Erlbaum.

WERBOS, P. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. Unpublished doctoral dissertation, Harvard University.

WIDROW, B., & HOFF, M. (1960). Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention Record*, **4**, 96-104.

WINOCUR, G. (1990). Anterograde and retrograde amnesia in rats with dorsal hippocampal or dorsomedial thalamic lesions. *Behavioral Brain Research*, **38**, 145-154.

WITTER, M. (1993). Organization of the entorhinal-hippocampal system: A review of current anatomical data. *Hippocampus*, **3**, 33-44.

WITTER, M., GROENEWEGEN, H., SILVA, F. L. D., & LOHMAN, A. (1989). Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region. *Progress in Neurobiology*, **33**, 161-253.

WOODHAMS, P., CELIO, M., ULFIG, N., & WITTER, M. (1993). Morphological and functional correlates of borders in the entorhinal cortex and hippocampus. *Hippocampus*, **3**, 313-312.

ZOLA-MORGAN, S., & SQUIRE, L. (1992). Components of the medial temporal lobe memory system. In L. Squire & N. Butters (Eds.), *Neuropsychology of memory* (2nd ed., pp. 325-335). New York: Guilford Press.

ZOLA-MORGAN, S., & SQUIRE, L. (1993). Neuroanatomy of memory. *Annual Review of Neuroscience*, **16**, 547-563.

ZOLA-MORGAN, S., SQUIRE, L., REMPEL, N., CLOWER, R., & AMARAL, D. (1992). Enduring memory impairment in monkeys after ischemic damage to the hippocampus. *Journal of Neuroscience*, **12**, 2582-2596.

NOTES

1. An alternate control condition for latent inhibition is to give an equal amount of "sit" time (context-only preexposure). We have simulated this by comparing simulations that receive 100 blocks of unreinforced exposure to A in context X with simulations that receive 100 blocks of unreinforced exposure to X alone. This is followed in both groups by 100 blocks of training to respond to A in X. In this case, the "sit" group gives an average conditioned response to A of 0.82 (SD 0.25), while the preexposed group gives an average response of only 0.56 (SD 0.31). This is a highly significant difference [$t(18) = 1.961, p < .05$], confirming the latent inhibition effect in the model.

2. We thank one of the anonymous reviewers of this paper for making us aware of this behavioral result.

APPENDIX
Simulation Details

This appendix describes the H-lesioned model proposed in this paper and shown in Figure 6. (The intact and H^{++}-lesioned models and the data plotted in Figure 2 are described in Gluck & Myers, 1993, and Myers & Gluck, 1994.)

Stimuli

Input to the system consists of a 16-element vector; the first 8 elements signal the presence or absence of conditioned stimuli (CSs), while the last 8 elements are nonvarying and represent background or contextual cues. The actual stimuli used in the experiments reported here are given below.

context X (no CS): 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0

CS A (in X): 1.0 1.0 1.0 1.0 0.0 0.0 0.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0

CS B (in X): 0.0 0.0 0.0 0.0 1.0 1.0 1.0 1.0 1.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0

CS AB (in X): 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0

A context shift is implemented simply by inverting the context bits:

context Y (no CS): 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0 1.0

CS A (in Y): 1.0 1.0 1.0 1.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0 1.0

For the easy task, the following stimuli are used:

    CS E+:   0.9 0.9 0.9 0.9 0.1 0.1 0.1 0.1 1.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0

    CS E−:   0.1 0.1 0.1 0.1 0.9 0.9 0.9 0.9 1.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0

For the hard task, the following stimuli are used:

    CS H+:   0.55 0.55 0.55 0.55 0.45 0.45 0.45 0.45 1.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0

    CS H−:   0.45 0.45 0.45 0.45 0.55 0.55 0.55 0.55 1.0 0.0 1.0 0.0 1.0 0.0 1.0 0.0

Criterion for learning was reached when for 10 consecutive epochs, the system response was greater than 0.8 for all reinforced trials and less than 0.2 for all nonreinforced trials. The Diff measure reported in figures describing discrimination tasks is the normalized difference between the average response to the positive and negative stimuli. All simulation data shown in Figures 7–10 are averaged over 10 randomly initialized simulation runs.

Entorhinal Network

The entorhinal network is adapted from the piriform clustering model (Ambros-Ingerson et al., 1990; Granger et al., 1989) and consists of a single layer of 100 nodes, divided into five nonoverlapping patches of 20 nodes each. Nodes in one patch are all reciprocally connected with a single local inhibitory feedback cell. As input is presented to the network, each node, n, calculates its output, y_n, as:

$$y_n = \sum_i w_{in} I_i,$$

where I_i is the ith element of the input vector and w_{in} is the weight from that element to node n of the network. After outputs are computed, the "winner" of each patch is selected as the node with maximum output, y_n. The feedback node is assumed to inhibit and normalize output so that, within each patch, the winning node outputs 1.0 while all other nodes output 0.0. The winning nodes then update their weights as

$$\Delta w_{in} = \alpha^+ I_i (1.0 - y_n).$$

Every node that is not a winner updates its weights as

$$\Delta w_{in} = \alpha^- I_i (0.0 - y_n),$$

where $\alpha^+ = 0.001$ and $\alpha^- = 0.0001$.

The entorhinal network weights, w_{in}, are initialized randomly with a uniform distribution and normalized so that for each node n,

$$\sum_i w_{in} = 1.$$

The original piriform model also assumes repetitive sampling and additional feedback inhibition to mask input features, which allow hierarchical cluster formation. These additional assumptions are not included in the entorhinal model as it is not clear whether such functional connections exist between the entorhinal cortex and its input structures.

Long-Term Memory Network

The long-term-memory network is adapted from the long-term network of the original hippocampal-region model presented in Gluck and Myers (1993) and represents a simplification of the cerebellar model presented in Gluck et al. (1994). Stimulus inputs feed into an internal or hidden layer of 10 nodes, which each compute output y_j as:

$$y_j = f\left(\sum_i w_{ij} I_i\right) + \theta_j,$$

where I_i is the ith element of the input vector, w_{ij} is the weight from that element to hidden node j, θ_j is the bias of node j, and $f(x) = 1/(1 + e^{-x})$. The output of these hidden nodes feed into a single output node k, which computes its output y_k as:

$$y_k = f\left(\sum_j w_{jk} y_j\right) + \theta_k,$$

where y_j is the output of hidden node j, $f(x)$ is defined as above, and θ_k is the bias of output node k. y_k is treated as the behavioral response of the system, and interpreted as the strength or probability of a CR on the current trial.

The long-term-memory network is trained by independent applications of the LMS learning rule (Widrow & Hoff, 1960) on each of its layers. The upper layer of weights w_{jk} are updated as:

$$\Delta w_{jk} = \beta(\lambda - y_k)y_j,$$

where λ represents the strength or salience of the US: $\lambda = 1$ if US is present and $\lambda = 0$ otherwise; y_j is the output of hidden node j and learning rate $\beta = 0.5$ if $\lambda = 1$ and $\beta = 0.05$ otherwise. The bias θ_k is trained as if it were a weight from a hidden unit which always outputs 1.0.

Simultaneously, the lower layer of weights w_{ij} are trained as:

$$\Delta w_{ij} = \beta(E_j - y_j)I_i,$$

where learning rate β is defined as above, and bias θ_j is trained as a weight from an input which always outputs 1.0. E_j is an aggregate training signal from the entorhinal network defined as:

$$E_j = \sum_n v_{nj}y_n,$$

where y_n is the output of entorhinal network node n, and v_{nj} is the connection strength from entorhinal node n to hidden node j in the long-term-memory network.

The lower layer of weights in the long-term-memory network are initialized from a uniform distribution in the range $[-0.3 \ldots +0.3]$, with added variance provided by initializing two weights per hidden unit in the range $[-3 \ldots +3]$. The upper layer of weights and all biases are also initialized from a uniform distribution in the range $[-0.3 \ldots +0.3]$. The connection strengths from the entorhinal network to the long-term-memory network are initialized to 0.0, with a random two connections per long-term-memory network hidden node initialized in the range $[-0.3 \ldots +0.3]$ and normalized to sum to 1.

Training Schedules

At the start of each experiment, the system is initialized by running for 500 trials with a null input vector and $\lambda = 0$. (The purpose of this initialization is to establish system weights and biases such that there is an initial tendency not to respond to novel vectors; before such initialization, system responses are about 0.5 for a novel vector.) Subsequent training occurs in epochs or blocks of trials. Each epoch consists of one presentation of each CS, and additional context-alone trials in a ratio of 20:1 with CSs. So, for example, A+ training would involve epochs of 10 context-only trials, 1 A+ trial, and 10 more context-only trials; A+,B− epochs involve 10 context-only trials, 1 A+ trial, 20 context-only trials, 1 B− trial, and 10 more context-only trials. Reinforcement (e.g., on A+ trials) is signaled by setting $\lambda = 1$; otherwise $\lambda = 0$.