

Cortico-Hippocampal Representations in Simultaneous Odor Discrimination: A Computational Interpretation of Eichenbaum, Mathews, and Cohen (1989)

Catherine E. Myers and Mark A. Gluck
Rutgers University

A previous model of hippocampal region function in classical conditioning is generalized to H. Eichenbaum, A. Fagan, P. Mathews, and N. J. Cohen's (1989) and H. Eichenbaum, A. Fagan, and N. J. Cohen's (1989) simultaneous odor discrimination studies in rats. The model assumes that the hippocampal region forms new stimulus representations that compress redundant information while differentiating predictive information; the piriform (olfactory) cortex meanwhile clusters similar and co-occurring odors. Hippocampal damage interrupts the ability to differentiate odor representations, while leaving piriform-mediated odor clustering unchecked. The result is a net tendency to overcompress in the lesioned model. Behavior in the model is very similar to that of the rats, including lesion deficits, facilitation of successively learned tasks, and transfer performance. The computational mechanisms underlying model performance are consistent with the qualitative interpretations suggested by Eichenbaum et al. to explain their empirical data.

Many forms of learning and memory are disrupted by hippocampal region damage (cf. Squire, 1987), including simultaneous odor discrimination in rats (Eichenbaum, Fagan, Mathews, & Cohen, 1988). The data suggest that hippocampal-lesioned rats tend to overcompress or "fuse" representations of co-occurring odors and are thus unable to respond selectively to the individual odors (Eichenbaum, Otto, & Cohen, 1992). However, occasionally, and seemingly at random, these lesioned rats can solve a discrimination as quickly as control rats (Eichenbaum et al., 1988). Intriguingly, these successful lesioned rats appear to use different strategies than control rats, as suggested by their abnormally poor performance on a subsequent transfer task involving novel pairings of previously learned odors (Eichenbaum, Mathews, & Cohen, 1989). Eichenbaum et al. (1989) interpreted these data as supporting their view that the hippocampus is needed for learning relations between stimuli that can be used flexibly in new contexts (cf. Bunsey & Eichenbaum, 1993; Eichenbaum et al., 1988). This is an important example of a growing body of recent data demonstrating that hippocampal-lesioned subjects may often show superficially normal performance but will differ from normal animals when challenged to use these learned associations in a novel way (cf. Saunders & Weiskrantz, 1989; Solomon & Moore, 1975; Winocur & Olds, 1978).

While a qualitative account based on flexibility provides an intuitive interpretation of these and other empirical data, it does not directly address the issues of how the hippocampus

performs this relational learning or how the learning might be transferred to other brain regions that are the presumed sites of long-term memory. It would be useful to have a more formal mechanistic characterization of what Eichenbaum refers to as *relational* learning that could predict a priori whether a particular task is expected to require such relational learning and, therefore, be sensitive to hippocampal region damage.

In Gluck and Myers (1993), we presented a computational model of hippocampal region function in classical conditioning; this model bears a strong conceptual similarity to the qualitative account of Eichenbaum and colleagues. In this cortico-hippocampal model, the hippocampal region is assumed to be critically involved in the formation of new stimulus representations that reflect both stimulus-stimulus and stimulus-outcome regularities in the environment. Hippocampal region damage is modeled as the loss of the ability to form new stimulus representations, although stimulus-response learning can still occur based on preexisting representations. To date, the model has focused on trial-level aspects of classical conditioning. Within this domain, the model has successfully accounted for a wide range of behavioral phenomena seen in both intact and lesioned animals and has made several novel predictions that remain to be tested (Gluck & Myers, 1993; Myers & Gluck, 1994).

Although it is possible to draw substantial parallels between this computational model and the more qualitative theories of Eichenbaum and colleagues (e.g., Eichenbaum & Bunsey, 1995; Gluck, Myers, & Goebel, 1994), the comparison has been hampered by the model's limited scope of applicability, viz, classical Pavlovian conditioning paradigms.

To enable a closer comparison between these two theoretical approaches to hippocampal function, we show that our computational model can be generalized to apply to the simultaneous odor discrimination task, as described by Eichenbaum et al. (1988, 1989). To accomplish this, we extend the

Catherine E. Myers and Mark A. Gluck, Center for Molecular and Behavioral Neuroscience, Rutgers University.

Correspondence concerning this article should be addressed to Catherine E. Myers, Center for Molecular and Behavioral Neuroscience, 197 University Avenue, Rutgers University, Newark, New Jersey 07102. Electronic mail may be sent via Internet to myers@pavlov.rutgers.edu.

model in two ways. First, we include a module representing olfactory preprocessing; this is based on an earlier model of stimulus clustering in piriform (olfactory) cortex (Ambros-Ingerson, Granger, & Lynch, 1990). Second, we consider training regimens in which reinforcement is contingent upon the behavioral response; this allows the model to instantiate forced-choice discrimination. With these straightforward extensions, the model is sufficient to account for many of the behavioral effects shown in control rats on this task. A hippocampal-lesioned version of the extended model captures many important aspects of the behavior of fornix-lesioned rats. The model further accounts for data showing that control, but not fornix-lesioned, rats perform well on a transfer task involving novel mispairings of trained discriminations.

Thus, the cortico-hippocampal model provides a computational framework for understanding Eichenbaum et al.'s empirical results. This suggests that these data reflect the same underlying representational processes that we previously proposed to occur in the hippocampal region during classical conditioning (Gluck & Myers, 1993; Myers & Gluck, 1994). These processes are related to the qualitative interpretations suggested by Eichenbaum et al. (1989). Furthermore, the fact that the hippocampal region model can, with minimal extensions, be applied to both classical and operant procedures supports the view that similar hippocampal-mediated processes underlie both kinds of learning.

A Computational Model of Hippocampal Region Function in Simultaneous Odor Discrimination

In this section, we first review our existing cortico-hippocampal model, and then we show how it can be generalized to apply to the simultaneous odor discrimination paradigm studied by Eichenbaum and colleagues. The essential extensions to the model are, first, to include piriform-cortex preprocessing of olfactory stimuli, and, second, to allow for a forced choice between multiple possible behavioral responses that in turn determine whether reinforcement arrives. The

resulting generalized model is sufficient to account for many of the effects observed by Eichenbaum and colleagues in their intact and lesioned rats.

Background: Classical Conditioning in the Cortico-Hippocampal Model

Gluck and Myers (1993) previously proposed that the hippocampal region is involved in the formation of new stimulus representations during learning. These new stimulus representations are sensitive to stimulus-stimulus regularities, the reliability with which two stimuli co-occur, and also stimulus-outcome regularities, the reliability with which one stimulus predicts that a second is about to occur. The representations are assumed to compress together stimuli that co-occur or predict similar outcomes, while differentiating stimuli that never co-occur or predict different outcomes. This representational learning is distinct from learning to associate a stimulus with a particular response; such stimulus-response learning is not assumed to depend on hippocampal region mediation, although it may be improved by utilizing the representations formed in the hippocampal region.

These processes can be implemented in a connectionist model, as shown in Figure 1 (Gluck & Myers, 1993). One network, a predictive autoencoder (Hinton, 1989), represents processing in the hippocampal region. External stimulus inputs activate an internal layer of nodes through weighted connections. The weight from a particular input to a particular internal layer node is a measure of how much that input affects the activity at that internal layer node. The entire pattern of internal layer node activations is a recoding or re-representation of the inputs. Internal layer node activity feeds through a second layer of weighted connections to activate nodes in the output layer. The network learns to map from its input pattern to outputs that reconstruct the input as well as predicting future reinforcement. This learning is accomplished by adjusting both the upper and lower layers of weights. Such adjustments can be calculated by a multilayer learning algorithm,

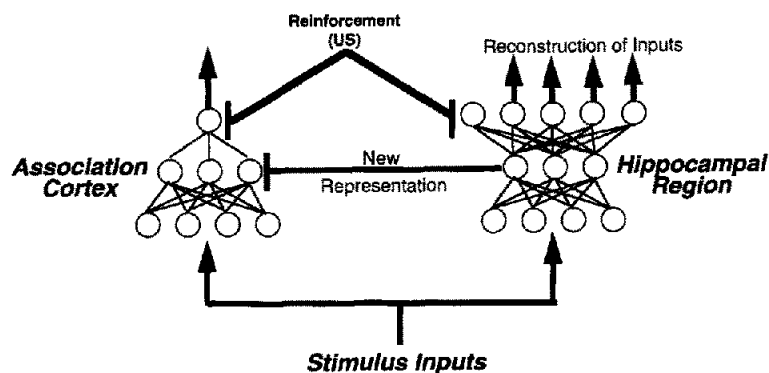


Figure 1. The cortico-hippocampal model presented in Gluck and Myers (1993). A hippocampal region network forms new internal representations that compress and differentiate information to reflect stimulus-stimulus and stimulus-outcome correlations. These representations are acquired by a second network that outputs the behavioral response and is the site of long-term memory. US = unconditioned stimulus.

such as error back-propagation (Rumelhart, Hinton, & Williams, 1986). Because the internal layer is narrow with respect to 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 internal layer representations tend to cluster and differentiate information based on both stimulus-stimulus and stimulus-output regularities. It is important to note that this architecture and learning algorithm represents only one of potentially many ways in which to instantiate these representational processes (see Myers, Gluck, & Granger, 1995); we make no particular claim here that this particular instantiation is the same as that used in the brain substrates, only that it produces appropriate functionality.

As originally specified (Gluck & Myers, 1993), the model 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-response learning but are assumed unable to construct the kinds of new stimulus representations developed in the hippocampal region. The cortices can, however, acquire the representations formed in the hippocampal region. One such network, representing some aspects of association cortex, is shown on the left in Figure 1; it learns to map from inputs, through its internal layer, to an output that is used to generate the behavioral response. This network is trained by a simple correlational learning rule, such as 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 & Steward, 1983; Stanton & Sejnowski, 1989). Using this rule, the network can learn to map from its existing internal layer representations to a prediction of future reinforcement. Using a second application of this rule, and sparse fixed connections from the hippocampal network, the association cortex network can also adopt a random linear recoding of the new representations evolving in the hippocampal network. Learning in both layers of the association cortex network and in the hippocampal region network is assumed to proceed incrementally and simultaneously.

As the association cortex acquires the hippocampal region representations in its internal layer, it finds a set of lower layer weights such that the same pattern (or a linear recombination of the pattern) that the current inputs activate in the hippocampal region network is activated in the association cortex. It is important to note that although the two networks may have the same patterns of internal layer node activity at this point, they may be generated by different combinations of lower layer weights. In general, the association cortex and hippocampal region probably do not have identical inputs, and so each makes use of whatever information it does receive in constructing a set of lower layer weights that activate an appropriate representation.

Within this framework, broad hippocampal region damage is simulated by disabling the hippocampal region network. The remaining association cortex network can no longer acquire new (hippocampal-mediated) representations. However, this

association cortex network can still train its upper layer of weights to map from existing (and now fixed) representations to new behavioral responses.

This model of learning has been applied to classical conditioning paradigms in which a response-evoking stimulus (the unconditioned stimulus or US) is repeatedly preceded by a neutral stimulus (the conditioned stimulus or CS). Over time, an association develops between CS and US such that the CS alone can elicit an anticipatory response (the conditioned response or CR). The model captures many trial-level aspects of classically conditioned learning in intact animals, including basic acquisition, latent inhibition, compound preconditioning, sensory preconditioning, successive reversal facilitation, the overtraining reversal effect, easy-hard transfer, and the nonmonotonic development of generalization gradients (Gluck & Myers, 1993). The model also provides an interpretation of how contextual information influences learning and correctly exhibits a response decrement with context shift, release from latent inhibition with context shift, and the ability of contextual cues to develop occasion-setting properties (Myers & Gluck, 1994).

Extension to Simultaneous Odor Discrimination

A basic experimental paradigm for forced-choice learning in rats is simultaneous odor discrimination, as described by Eichenbaum and colleagues (Eichenbaum et al., 1988, 1989). As shown in Figure 2, a rat is placed in a small chamber with a cul-de-sac containing two odor delivery ports. On each trial, two odors, A and B, are presented simultaneously, one from each port but with varying spatial location. One odor (e.g., A) is arbitrarily designated as positive, and the rat is given a water reward for poking and holding its nose in the port delivering odor A. On each trial, both stimuli are present, and so it is not enough simply to learn to respond to the presence of stimuli. The correct spatial response—nosepoke left or right—is determined by the spatial arrangement of the stimuli. For instance, in the A+B- discrimination, either stimulus may appear from the left odor port. Thus, there are actually two kinds of trial, AB (A, left and B, right) and BA (B, left and A, right). It is the relative locations of the stimuli on each trial that determine the correct response.

The rat is considered to have acquired the discrimination when it satisfies a criterion of 18 correct responses within a block of 20 trials; the discrimination is terminated if it is not acquired within 600 trials. Following acquisition or termination of the discrimination, the rat is trained on a new discrimination involving two new odors (e.g., C and D) but with identical task demands. Additional discriminations with more novel odor pairs may follow.

This simultaneous odor discrimination paradigm differs in several important ways from the classical conditioning tasks to which our cortico-hippocampal model has previously been applied (Gluck & Myers, 1993; Myers & Gluck, 1994), and the model must be extended to reflect these differences. The most important difference is that in the classical conditioning tasks, arrival of the reinforcer (US) is contingent only on presentation of the CS; this contingency is independent of any conditioned responding. The odor discrimination task, by contrast,

is an operant procedure, in which reinforcement is delivered contingent on the response.¹ Thus, the model must now include this contingency; further, because there is no way to know a priori whether other, unexecuted responses would have been effective, the learning procedures must be altered so that learning applies only to those response–reinforcement contingencies that are actually experienced.

A second important distinction is that we have previously modeled classical conditioning preparations in which the form of the response evoked by the reinforcer and CS are similar. For example, in conditioned eyeblink responding, the US is a blink-evoking airpuff, and after repeated CS–US pairings, the CS also comes to evoke a protective eyeblink. This conservation of response form does not apply in the odor discrimination task: here, the conditioned response is an approach to the positive odor port, which differs from the drinking response elicited by arrival of the reinforcer. Thus, the model must be extended to allow for the possibility of multiple behavioral responses, which are each different in form from the response evoked by the reinforcer. Finally, since the original model was applied to tasks involving only one or two conditioned stimuli, it was possible to use a simplified input format that did not take into account any stimulus preprocessing. For the odor discrimi-

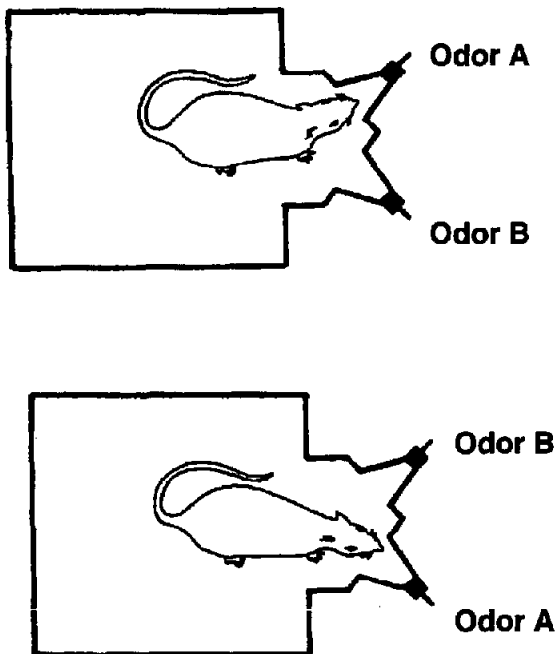


Figure 2. Simultaneous odor discrimination paradigm. The rat is presented with two odors, A and B, from two odor ports; the spatial arrangement of A and B varies across trials. One odor (e.g., A) is arbitrarily but consistently designated as positive, and the rat obtains a reward for a nosepoke to the odor port delivering A. *Note:* From “Hippocampal System Dysfunction and Odor Discrimination Learning in Rats: Impairment or Facilitation Depending on Representational Demands,” by H. Eichenbaum, A. Fagan, P. Mathews, and N. J. Cohen, 1988, *Behavioral Neuroscience*, 102(3), p. 331. Copyright 1988 by the American Psychological Association. Reprinted with permission of the author.

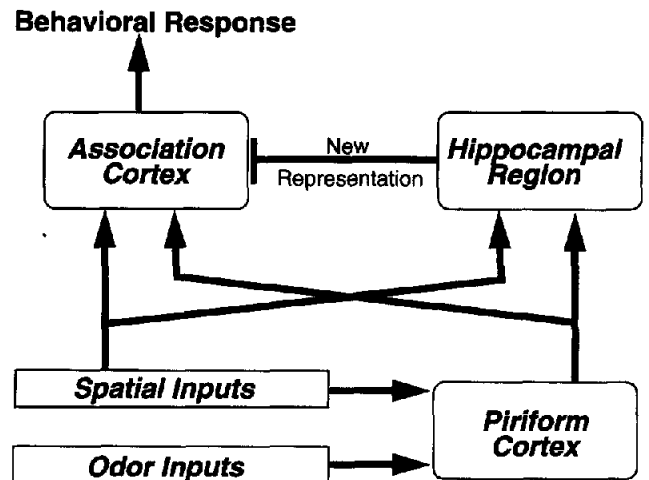


Figure 3. Schematic of information flow in the generalized cortico-hippocampal model. The association cortex receives both odor inputs and information about their spatial location; the hippocampal region receives the spatial information, as well as odor information preprocessed by the piriform cortex. The hippocampal region continues to provide new stimulus representations to the association cortex, and the association cortex output is used to generate the behavioral response.

nation task, there are multiple stimuli taken from the same modality, and, therefore, we have considered how these inputs might be preprocessed by the olfactory cortex.

These generalizations, and the resulting model, are described in the following.

Stimulus inputs. Figure 3 shows the information flow in the generalized version of the model. We assume that the external sensory input to the model details which odors are present and which odor ports they arise from. Because the stimuli used in this paradigm are explicitly chosen to be highly distinctive (cf. Eichenbaum et al., 1988), we have assumed nonoverlapping representations.

In addition to this odor identification information, the discrimination paradigm also requires use of spatial information. On each trial in the discrimination task, there is exactly one odor present at each of the two spatial (left and right) locations. The process by which specific odors are bound to locations is beyond the scope of the current model, and so we have simply assumed an array in which a unique element codes for the spatial location of each stimulus. This spatial information provides additional input to both the association cortex and hippocampal region networks.

¹ Although this is an operant procedure, it is possible that animals may solve it through use of classical rather than operant strategies; that is, they may simply learn a classically conditioned approach response to the positive odor (or avoidance of the negative odor or both). Further empirical studies would be needed to fully resolve this issue. However, our primary interest here is in extending our model to apply to a particular operant procedure and to examine hippocampal-mediated stimulus representation during this task. These processes should be largely the same whether the animal is using classical or operant strategies to solve the task.

Odor preprocessing. In our modeling work to date, we have assumed that input is preprocessed before it reaches the hippocampal region, but we have ignored the details of how that preprocessing might occur. Because the model as originally specified focused on classical conditioning studies involving a small number of highly distinct stimuli drawn from different modalities, within-modality preprocessing was presumed to be only of secondary importance. Now, however, if the model is to be applied to simultaneous discrimination among large numbers of olfactory stimuli, the within-modality preprocessing of these stimuli is potentially far more germane. For this reason, we extend the model to include a module representing olfactory clustering in the primary olfactory (piriform) cortex (Ambros-Ingerson et al., 1990). This module is a simplification of a bottom-up, physiologically motivated model of piriform cortex, and it functions to construct stimulus representations in much the same spirit as does the hippocampal region network.

In brief, Ambros-Ingerson et al. (1990) modeled the superficial piriform cortex as a competitive learning system, with sparse afferent connections to target cells and denser local inhibitory feedback connections (see Figure 4). As the target cells are activated, they in turn activate the inhibitory cells, which suppress all but the most strongly activated target cells. These target cells undergo synaptic plasticity to increase their likelihood of winning the competition again when similar inputs are presented in the future. This network is assumed to be organized into nonoverlapping clusters, each with a single inhibitory feedback cell interacting with several target cells, only one of which can win a given competition.²

One emergent property of the operation of the piriform network is stimulus clustering. Stimuli with high superficial similarity will tend to generate the same response patterns across the network. That is, if a target cell wins the competition for a given stimulus pattern, and undergoes plasticity to make it more likely to win the competition for that stimulus in the future, then it is also more likely to win the competition to respond to similar stimuli in the future.

A second emergent property of the piriform network is redundancy compression. If two stimuli co-occur, they will be treated as a single complex stimulus by the network, and a cluster will be formed in response to the features of this complex stimulus. If one of the component stimuli later occurs alone, the network will treat this as a degraded version of the compound stimulus and will respond in kind by retrieving the cluster associated with the compound. This is equivalent to the stimulus-stimulus redundancy compression function defined earlier in the context of the proposed hippocampal region function.

We have elsewhere noted the correspondence between the anatomy, physiology, and plasticity of the superficial piriform cortex and the neighboring entorhinal cortex, one structure of the hippocampal region (Myers et al., 1995). We have proposed that whereas clustering and compression of olfactory stimuli can emerge from the piriform cortex, similar function could arise from the entorhinal cortex. Because the entorhinal cortex receives multimodal and cross-modal inputs, as well as inputs from piriform cortex, the entorhinal cortex might be

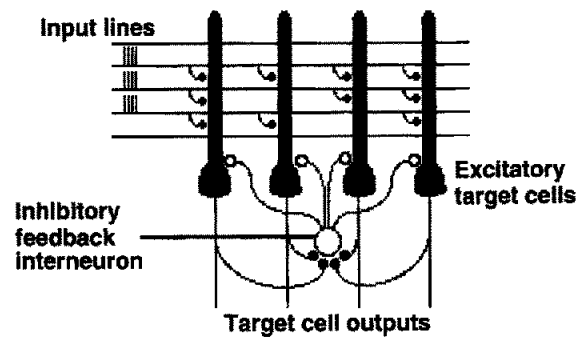


Figure 4. A simplified version of the piriform-clustering model (Ambros-Ingerson et al., 1990). Target cells are excited by sparse afferents, and they in turn activate local inhibitory feedback interneurons. Feedback silences all but the most strongly activated target cells; synaptic plasticity makes these "winning" target cells more likely to win in response to similar inputs in the future. *Note.* From "Dissociation of Hippocampal and Entorhinal Function in Associative Learning: A Computational Approach," by C. Myers, M. Gluck, and R. Granger, 1995, *Psychobiology*, 23(2), p. 116. Copyright 1995 by the Psychonomic Society, Inc. Reprinted with permission.

well placed to perform clustering and compression across modalities and between the multimodal features of a single stimulus. This hypothesis has important implications for the effects of entorhinal damage on compression on tasks within and between modalities. We will return to this issue later, in the conclusions section.

For now, we note that most cortical input to the hippocampal region arrives via the entorhinal cortex (Witter, 1993), as schematized in Figure 5. The entorhinal cortex receives inputs from multimodal association cortices, such as the (primarily visual) perirhinal cortex and (primarily visuospatial) parahippocampal cortex; it also receives a strong input from the piriform cortex (Suzuki, 1994). Therefore, we assume that input to the cortico-hippocampal model should include the output from the piriform cortex model, which performs stimulus clustering of odor inputs, as well as spatial information detailing the locations of odor stimuli. These same inputs detailing odor features and locations are provided as input to the association cortex module as well.

Simultaneous odor discrimination. To apply to the simultaneous odor discrimination, the model must choose among the possible behavioral responses (here, nosepoke left or right) and then receive reinforcement based on whether the selected response was correct. This is done by allowing the association

² The full piriform model described by Ambros-Ingerson et al. (1990) also assumes repetitive sampling of stimulus inputs, with recurrent feedback from the piriform cortex to olfactory bulb masking the input patterns on subsequent samples. This repetitive sampling allows hierarchical clustering of stimulus inputs. We have not assumed this repetitive sampling in our simplified version of the piriform model, nor the hierarchical clustering, and will not discuss these further here. Thus, this reduced version of the piriform model is simply a competitive learning system analogous to those described by Grossberg (1976), Kohonen (1984), Rumelhart and Zipser (1985), and others.

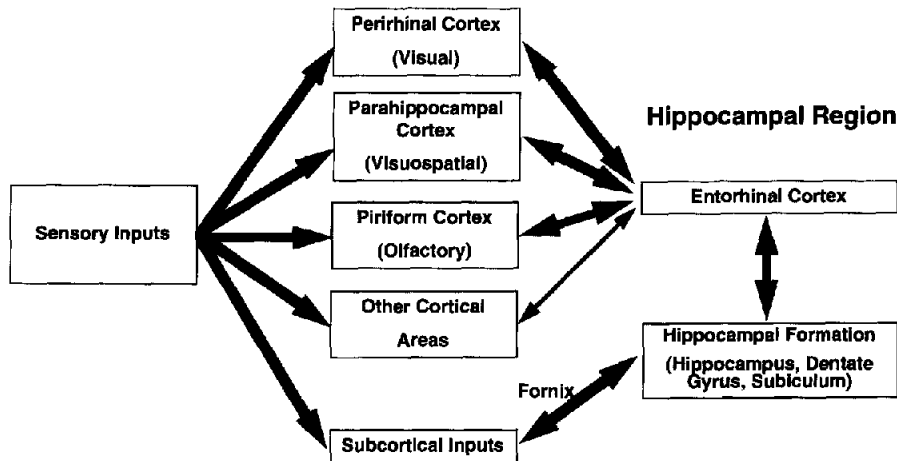


Figure 5. The hippocampal region receives inputs from a range of multimodal and association cortices, including the (primarily visual) perirhinal cortex, the (primarily visuospatial) parahippocampal cortex, and the (olfactory) piriform cortex.

cortex output to have two output nodes, each representing one of the possible spatial responses. These possible responses are then translated into a behavioral response, presumably through the mediation of other brain structures such as motor cortex and cerebellum. For the purposes of the model, we simply assume that the output activations from association cortex pass through a response generator that converts them to normalized response probabilities via a ratio response rule (Hull, 1943; Luce, 1963; Shepard, 1957; Thurstone, 1927) and then generates a behavioral response: poke left or poke right.

The selected response is then compared with the location of the positive stimulus, and a binary reinforcement signal reflects whether they match. This reinforcement is used only to train the output node corresponding to the chosen response, because it does not necessarily indicate whether alternate responses would have been rewarded or not.

The hippocampal region network continues to learn to reconstruct its inputs, but it additionally learns to predict the behavioral response. Thus, it also has multiple output nodes, one for each possible behavioral response. As always, the internal representations formed in the hippocampal region network are used to train the internal representations in the association cortex. The generalized cortico-hippocampal model is shown in Figure 6A, and full implementation details are given in the Appendix.

The lesioned model. Eichenbaum et al. (1988, 1989) compared the performance of control rats to rats with fornix transection. The fornix is a fiber pathway that connects the hippocampus with subcortical structures such as the thalamus and septum (Swanson, 1979). The fornix appears necessary for proper hippocampal function (cf. Port & Patterson, 1984; Saunders & Weiskrantz, 1989). Researchers have often argued that fornix lesion impairs learning as severely as outright hippocampal removal (e.g., Otto, Schottler, Staubli, Eichenbaum, & Lynch, 1991). However, rather than explicitly destroying hippocampus, fornix lesion presumably interrupts a critical modulatory pathway (Hasselmo & Schnell, 1994); for instance, the theta rhythm in hippocampus, which has been linked to

learning (Berry & Thompson, 1979; Buzsaki, 1989; Solomon & Gottfried, 1981), appears to be modulated by septum via fornix. Therefore, there may be important differences between fornix disruption and hippocampal removal (cf. Zola-Morgan, Squire, & Amaral, 1989b). We have argued elsewhere that while the effects of outright hippocampal lesion may compare to removing the hippocampal region network in the model, the effects of hippocampal disruption may be more comparable to reducing the rate at which new information is stored in the hippocampal network, though not the rate at which information is transferred from hippocampus to cortex (Myers et al., 1996).

Accordingly, we consider the lesioned version of the generalized cortico-hippocampal model seen in Figure 6B. The hippocampal region network is assumed to be unable to learn any new representations, though it continues to produce output that is a random, fixed remapping of its inputs. The rest of the system is otherwise unchanged from the intact model of Figure 6A.

Because the hippocampal region network no longer adapts its representations, its outputs to the association cortex network are no longer biased by predictive differentiation or by redundancy compression across both odor information and spatial location. Instead, they are dominated by the odor clusterings developed by the piriform network and also the raw spatial input. The piriform outputs will tend to emphasize superficial similarity of odor inputs and to de-emphasize spatial location. They will also tend to compress together the representations of co-occurring odors, as discussed earlier. In the case of a simultaneous discrimination, this compression will impair learning to respond based on the spatial locations of those odors. In rare cases, where the piriform network output preserves spatial information, or where the hippocampal representation happens to emphasize its spatial inputs, learning a simultaneous odor discrimination may proceed unchecked. However, in general, the result will be overcompressed odor representations that do not maintain spatial information; this overcompression in the lesioned model gives

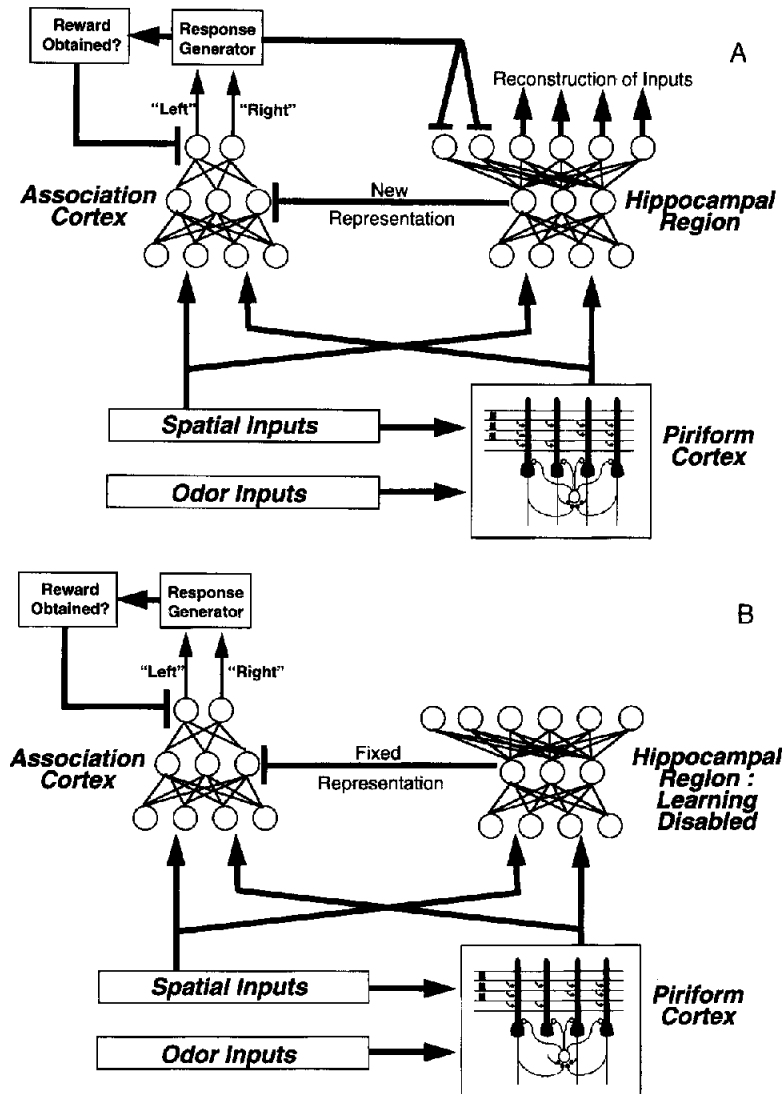


Figure 6. The cortico-hippocampal model is extended to allow olfactory preprocessing as well as odor discrimination learning. A: The intact model. The association cortex network now has two outputs, and a response generator chooses between them according to a probabilistic rule. Reward is contingent on the correct response. The hippocampal region network is also adapted to output a reconstruction of its inputs plus a prediction of the behavioral response. B: The fornix-lesioned model in which the hippocampal region network is disabled and can no longer construct new stimulus representations. The output of the hippocampal region internal layer is thus a fixed, random transform of the hippocampal network inputs: the spatial inputs and the compressed representations output by the piriform cortex network.

rise to many of the behaviors seen in fornix-lesioned rats on the simultaneous odor discrimination task, as discussed next.

**Simultaneous Odor Discrimination:
Empirical and Model Data**

In their studies of simultaneous odor discrimination, Eichenbaum et al. (1988, 1989) found several basic effects. These include a relative impairment in lesioned rats, progressive facilitation of successive discriminations in intact rats, bimodal distribution of solution times in the lesioned rats, and good transfer performance on mispaired odors in the intact but not

lesioned rats. We show that the generalized cortico-hippocampal model can capture each of these behaviors correctly and, in each case, does so for mechanistic reasons that are quite similar to the qualitative explanations previously offered by Eichenbaum and colleagues to explain their data.

Simultaneous Odor Discrimination in Rats and Model

In the basic simultaneous odor discrimination paradigm, rats are trained on several discriminations successively. As Figure 7A shows, control rats learn the first discrimination within a few hundred trials, and they learn subsequent discrimi-

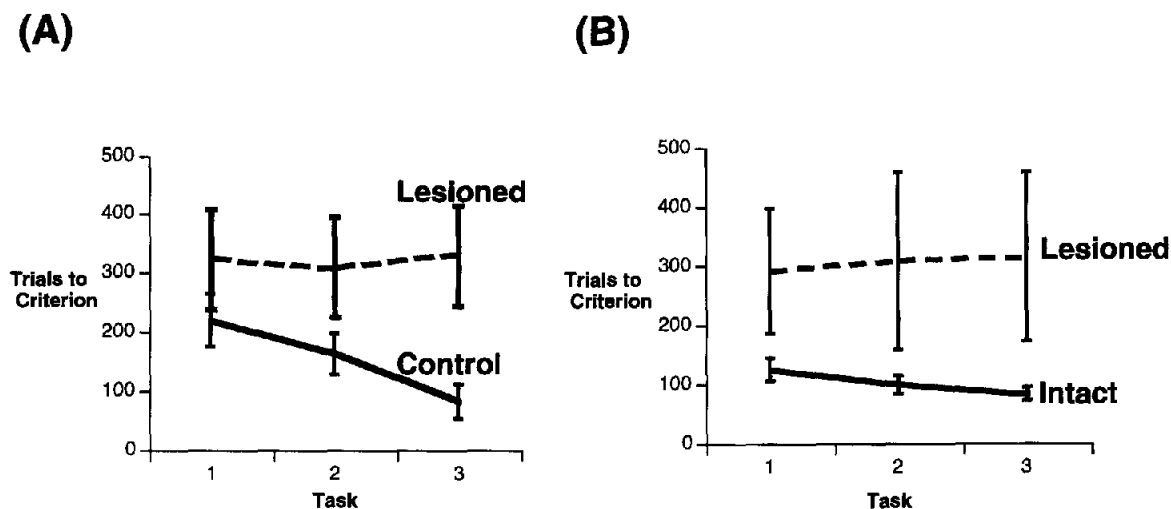


Figure 7. Simultaneous odor discrimination. A: Control rats show a facilitation on successive discriminations, requiring fewer training trials to reach criterion performance. Fornix-lesioned rats show no such facilitation. Figure is replotted from data presented in Eichenbaum et al. (1988). B: Like the rats, the intact but not lesioned model shows a significant facilitation on successive discriminations.

nations between new odor pairs progressively faster (Eichenbaum et al., 1988). By contrast, fornix-lesioned rats are strongly impaired at the task and show no facilitation with successive discriminations (Eichenbaum et al., 1988). On each of the three discriminations, about half of the lesioned rats fail to reach criterion performance.

Figure 7B shows that the intact and lesioned models show comparable performance to the control and lesioned rats. The intact model learns its first discrimination within an average of 124.4 training epochs, and later discriminations are learned faster; the third discrimination is learned within an average of 81.7 training epochs. The intact model's facilitation in learning the second discrimination relative to the first is highly significant, $t(9) = 3.44, p < .005$, and the facilitation of the second discrimination relative to the third is also highly significant, $t(9) = 2.19, p < .05$. The lesioned model shows a severe impairment on simultaneous discrimination and no facilitation on subsequent discriminations. The difference in learning times for the various discriminations all fail to reach significance ($p > .1$). Using a 300-trial pass-fail criterion similar to that used by Eichenbaum et al. (1988), we found that 40% of discriminations are not solved by the lesioned model, a failure rate comparable to that of the lesioned rats.

Why does the lesioned model show an impairment on this task, while the intact model does not? In the intact model, the piriform model clusters odor inputs based on similarity and co-occurrence. For a given discrimination, say A+B-, the two odors A and B always co-occur, and so the piriform model will tend to cluster their representations. This information, together with the spatial inputs, is passed to the hippocampal region network. The hippocampal region network forms new internal representations that, among other things, must differentiate the information needed to solve the task. In this case, that information is which odors appear from which ports—whether the current trial is an AB or BA trial. All other

information, including the odor cluster information arriving from the piriform cortex, is redundant for this end. (Note, however, that the hippocampal network is also required to reproduce all of its inputs, and so no input information can be completely ignored in the hippocampal network stimulus representations.)

The way that the hippocampal region network creates different representations for AB and BA is to make sure that different internal layer nodes are activated by these two kinds of trials. Figure 8A shows that the internal layer nodes in the hippocampal region network are activated by two kinds of inputs: those from the piriform network, and those detailing spatial information. The hippocampal region network can construct a representation that emphasizes the latter spatial information by selectively weighting connections from the inputs carrying that spatial information and by assigning relatively weaker weights to the inputs from the piriform network. These weights may be positive (excitatory) or negative (inhibitory); it is their magnitude that matters. When the hippocampal-mediated representation is adopted by the association cortex, it must find a set of lower layer weights that allow its internal layer nodes to produce the same representations. Thus, it also should generate strong weights for the inputs carrying spatial information and weaker weights for the inputs carrying compressed piriform outputs (see Figure 8B).

Figure 9 shows that the hippocampal region network starts with lower layer weights that are all random and generally small (see Figure 9A). After 500 epochs of training on the A+B- discrimination, exactly those weights that code for the spatial location of A and B (i.e., which is on the right and which is on the left) grow in strength (see Figure 9B). Thus, their ability to influence the internal representation is high. This reflects the hippocampal-dependent mechanism of predictive differentiation: Because the spatial locations of A and B are especially predictive of the correct behavioral response, these

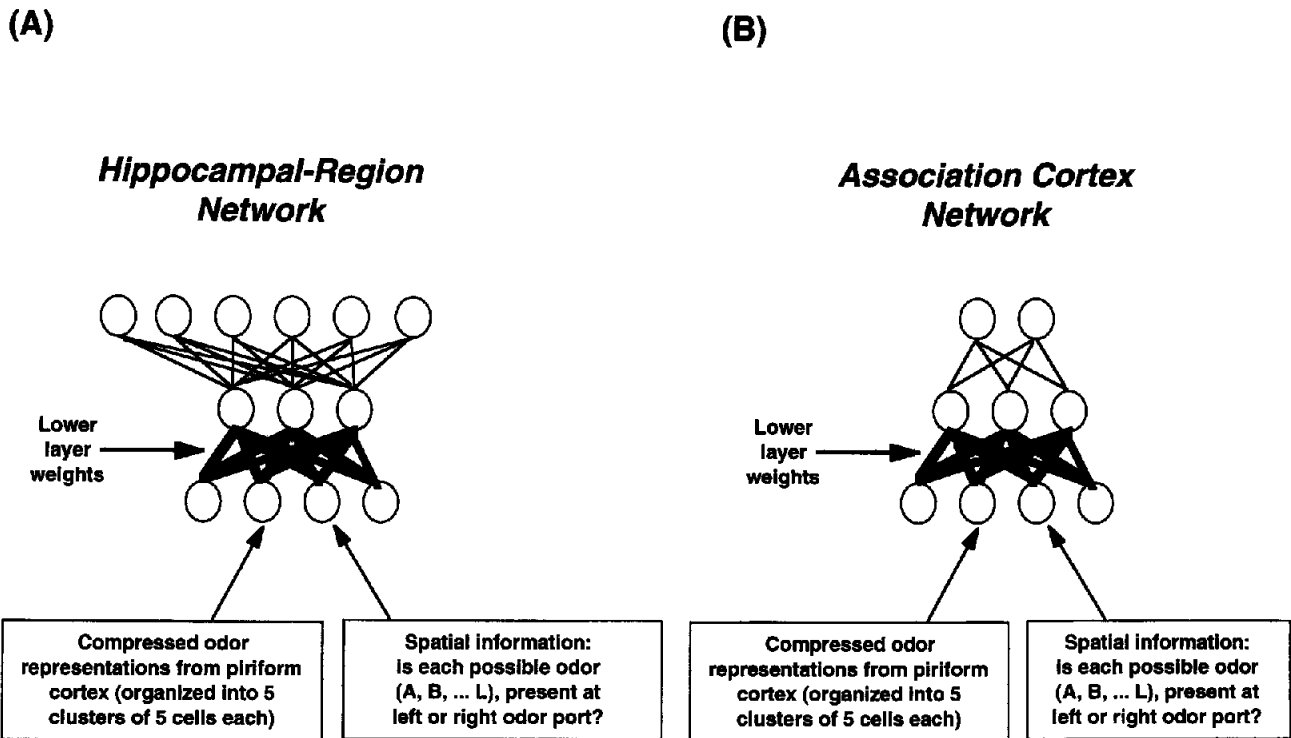


Figure 8. Stimulus representations in the internal layers of the intact model networks depend on the connections strengths (weights) from various inputs. A: The hippocampal region network receives inputs from the piriform cortex as well as spatial information; if the latter have strong weights, the internal layer representation will emphasize spatial information. B: The association cortex network receives the same two kinds of inputs; as it adopts the representations formed in the hippocampal region network, it must find lower layer weights that allow it to reconstruct those representations. If the hippocampal region representation emphasizes spatial information, the association cortex will selectively assign strong weights to inputs containing spatial information.

inputs are strongly weighted to allow maximally different representations in the internal layer depending on the placement of the two odors. A few of the weights from the olfactory inputs also grow; these are used by the hippocampal region network as additional information regarding which odor inputs are present. Although this is superfluous, and could be deduced from the spatial input alone, it is a property of the autoencoding hippocampal region network to include information about all relevant inputs in its internal layer representation. After 500 more epochs of training on the C+D-discrimination, these weights remain strong, but the weights encoding for the spatial location of C and D have now grown in strength (see Figure 9C). Again, the internal representations devote a great deal of emphasis to the spatial locations of the odor stimuli, and they largely de-emphasize all the other inputs. Later, in Figure 12, we will consider the contribution of individual internal layer nodes to the average profile shown in Figure 9D.

This pattern of weight changes contrasts with that observed in the lesioned model. In the lesioned model, the piriform network continues to perform odor clustering, but the hippocampal region network no longer can differentiate representations based on spatial information. Thus, the new representations acquired in the lesioned model's association cortex

network are based wholly on the piriform network output. Because the piriform network output tends to cluster co-occurring stimuli, most of the spatial information regarding their locations is not preserved. Figure 9E shows the mean absolute values of the lower layer of weights in a representative lesioned model's association cortex network after 500 epochs of training on each of the A+B- and C+D- discriminations; little if any differentiation of the cue locations has occurred. Because the cue location is exactly the information needed to solve the discriminations, the lesioned network shows a profound impairment at this task. Conceptually, the lesioned model treats A and B as components of a single compound stimulus, ignoring their spatial arrangement, and therefore it does not distinguish AB and BA trials. Eichenbaum et al. (1989) interpreted the lesioned-rat data in a similar way. They suggested that fornix-lesioned rats perceive stimulus pairs as unitary compounds; for each such compound, nose pokes to the right and left odor ports are each sometimes rewarded, and the rat cannot predict the correct response on a given trial.

On occasion, depending on the random initial configuration of weights in the lesioned network, some spatial information is preserved in the internal representation. Figure 9F shows the lower layer of weights for a different lesioned network trained

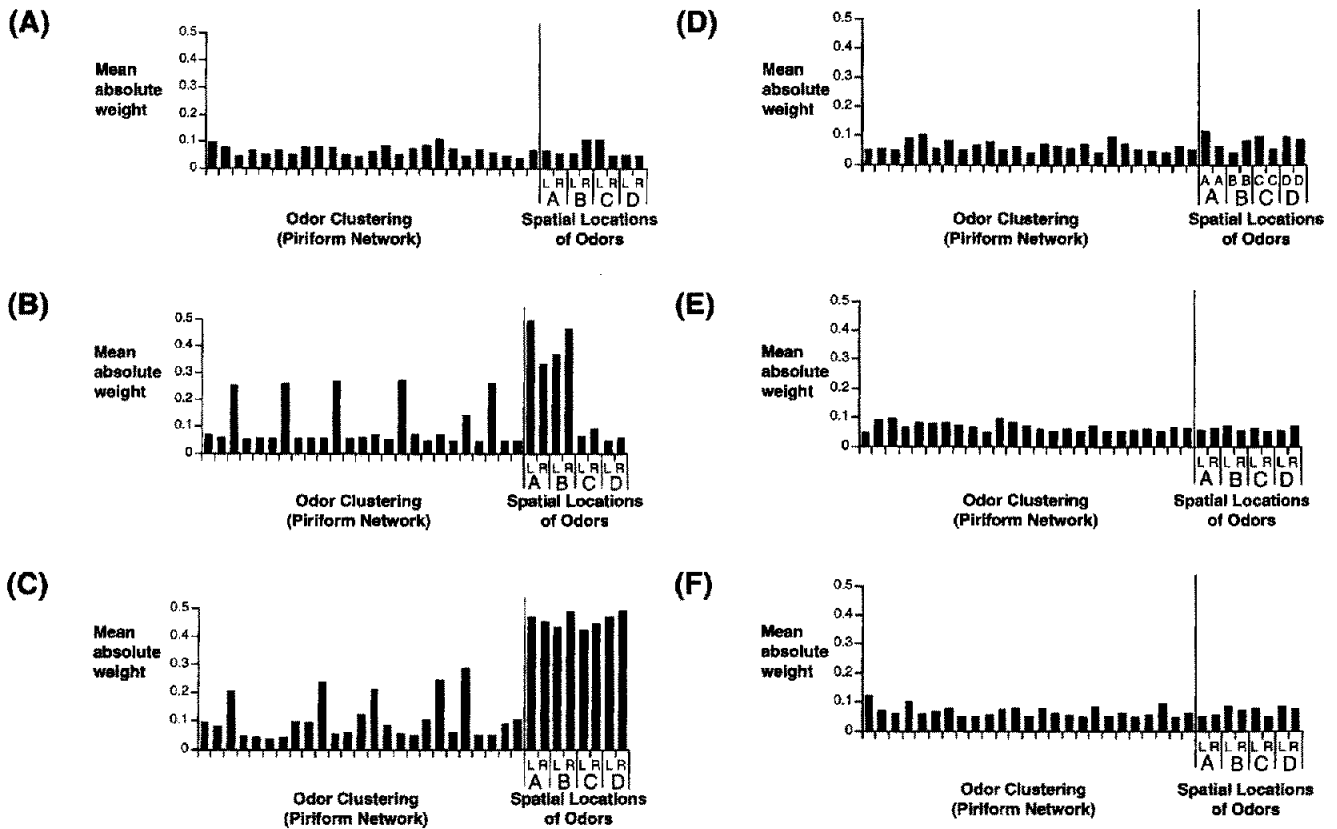


Figure 9. Magnitude of weights from the inputs to the internal layer nodes in the association cortex, which determine the internal representations activated, shown averaged across all internal layer nodes. For simplicity, only weights from the four relevant odors (A, B, C, and D) are shown. A: Initially, in a representative intact model, all weights are random and small. B: After training on A+B-, specifically those weights detailing the spatial locations of A and B gain strength. C: After additional training on C+D-, the weights from inputs indicating the spatial locations of C and D also increase; the weights indicating the locations of A and B remain strong. D: In a representative lesioned model, there is the same profile of initial weights. E: After training on A+B- and C+D-, there is little increase to the relevant spatial weights: as a result, neither task is learned by this network. F: Occasionally, a lesioned network does manage to solve both the A+B- and C+D- discriminations. However, even in such a case, the average weight profile still shows little consistent growth on the relevant spatial weights. Although one or two individual units may maintain some spatial information—allowing the model to solve the discrimination—there is no overall pattern as shown in the intact model in Figure 9C. L = left; R = right.

on the A+B- and C+D- discriminations; in contrast to Figure 9E, this network happens to preserve some spatial information. It will be noted that the magnitude of these weights is much smaller than that shown in Figure 9C for an intact network; nonetheless, this lesioned network does manage to solve the discriminations successfully. We will return to this point further, after the next section.

Facilitation of Successively Learned Discriminations in Intact Rats and Model

A second point of similarity between the animal and model data shown in Figure 7 is that both the control rats and intact network show a facilitation of subsequent discriminations after the first is learned. In the computational model, it is possible to see exactly why this facilitation is obtained. Figure 9, A and B,

shows how learning occurs in the intact model lower layer weights to differentiate the spatial locations of A and B. These changes take place first in the hippocampal region network and are then adopted by the association cortex, where they begin to influence the behavioral response. This transfer is not immediate, but it occurs with a slight time lag, as shown in Figure 10. Eventually, the same differentiation between AB and BA seen in the hippocampal region network internal representation is mirrored in the association cortex network internal representation. Now a second C+D- discrimination is trained; Figure 9, B and C, shows that only a few lower layer weights must undergo significant change, namely those encoding the spatial locations of C and D. The association cortex network must only acquire these few changes before the task can be solved; this is done more quickly than in the original

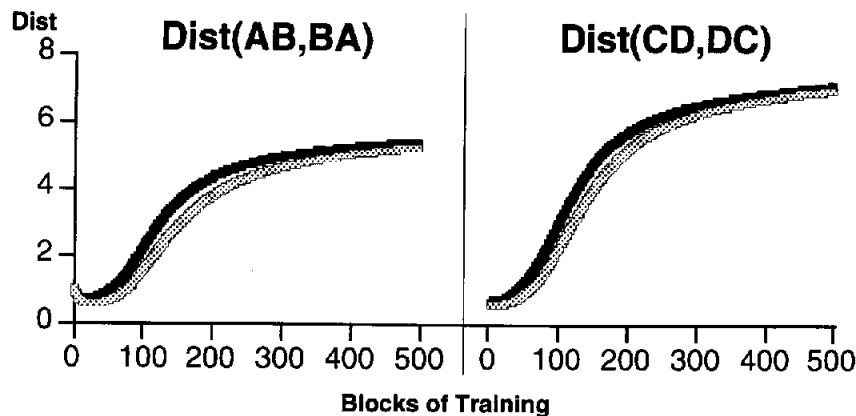


Figure 10. Differentiation of internal representations in a representative simulation, in terms of the summed difference in internal layer node activations to two input patterns, in the hippocampal network (black line) and association cortex network (gray line). During the first A + B- discrimination, the hippocampal region network develops representations that differentiate the two spatial arrangements of stimuli AB and BA. With a slight lag, the association cortex network adopts these representations, mirroring the differentiation of AB and BA. During the next C + D- discrimination, the hippocampal region develops representations that differentiate the two spatial arrangements CD and DC. Because there is a high overlap between the input patterns in the A+B- and C+D- discrimination, particularly in terms of the large number of irrelevant inputs, the hippocampal network differentiates the input patterns in this second discrimination more quickly. This in turn leads to a facilitation in learning the second C+D- discrimination, relative to the earlier A+B- discrimination. Dist = Hamming distance.

task (see Figure 10), and so there is a facilitation on the second task. Progressively more facilitation may be obtained with further discriminations, resulting in the pattern of results seen in Figure 7B. Eichenbaum et al. (1989) suggested that progressive facilitation of discrimination learning in the intact rats might reflect learning set acquisition; the modeling results suggest that it is not necessary to invoke higher level cognitive explanations to account for the facilitation: a representational account suffices. We note that this account is consistent with the finding that, during conditioning, cellular activity in the hippocampus mirrors and precedes the development of the behavioral response (Berger & Thompson, 1978). To establish whether this account is an accurate description of unit activity during the odor discrimination task would require electrophysiological studies monitoring the evolution of hippocampal cell activity during acquisition of the discriminations; it is certainly true that in the well-trained rat hippocampal cells exist that fire maximally during the sampling of specific, meaningful odors in particular spatial configurations (cf. Wiener, Paul, & Eichenbaum, 1988), which would be consistent with this account.

Bimodal Distribution of Solution Times in Lesioned Rats and Model

Curiously, while the fornix-lesioned rats are generally impaired on simultaneous discrimination, they occasionally solve some problems; when they do, they solve them as quickly as control rats, as shown in Figure 11, A and C.

The lesioned model shows a similar effect. Generally, the overcompressed representations in the lesioned model fail to preserve the spatial information needed to acquire the discrimi-

nation. On occasion, however, the representation may fortuitously preserve information regarding stimulus arrangement. If so, then that information may be simply mapped to the correct output response. In this case, learning may occur very quickly in the lesioned model. Combining the simulation data across all three discriminations, the lesioned model shows a bimodal distribution of learning time, as shown in Figure 11D: tasks are either unlearned or learned about as quickly as by the intact model. Using a 300-epoch pass-fail cutoff similar to that reported in Eichenbaum et al. (1988), the intact model reaches criterion on 30 discriminations and fails on none (see Figure 11B), while the lesioned model succeeds on 18 and fails on 12. A chi-square analysis of this distribution indicates a highly significant difference between the intact and lesioned model performances, $\chi^2(1, N = 30) = 15.0, p < .01$. In this way, the model correctly accounts for the different distributions of learning rates in the intact and lesioned rats.

Why do some lesioned systems learn some discriminations quickly? To answer this question, it is necessary to look at the roles of individual internal layer nodes in the model. Because the internal representation as a whole is a concatenation of the activity levels of these units, saying that two stimuli evoke broadly different internal representations means that those stimuli evoke broadly different activity levels on at least some of the internal units. A particular unit will respond differently to two stimuli, say AB and BA, if it has strong weights from the inputs detailing the locations of A and B and if these weights differ in magnitude for the two spatial arrangements of these stimuli.

For example, Figure 12 shows the weights to each internal layer node in the association cortex network of an intact system

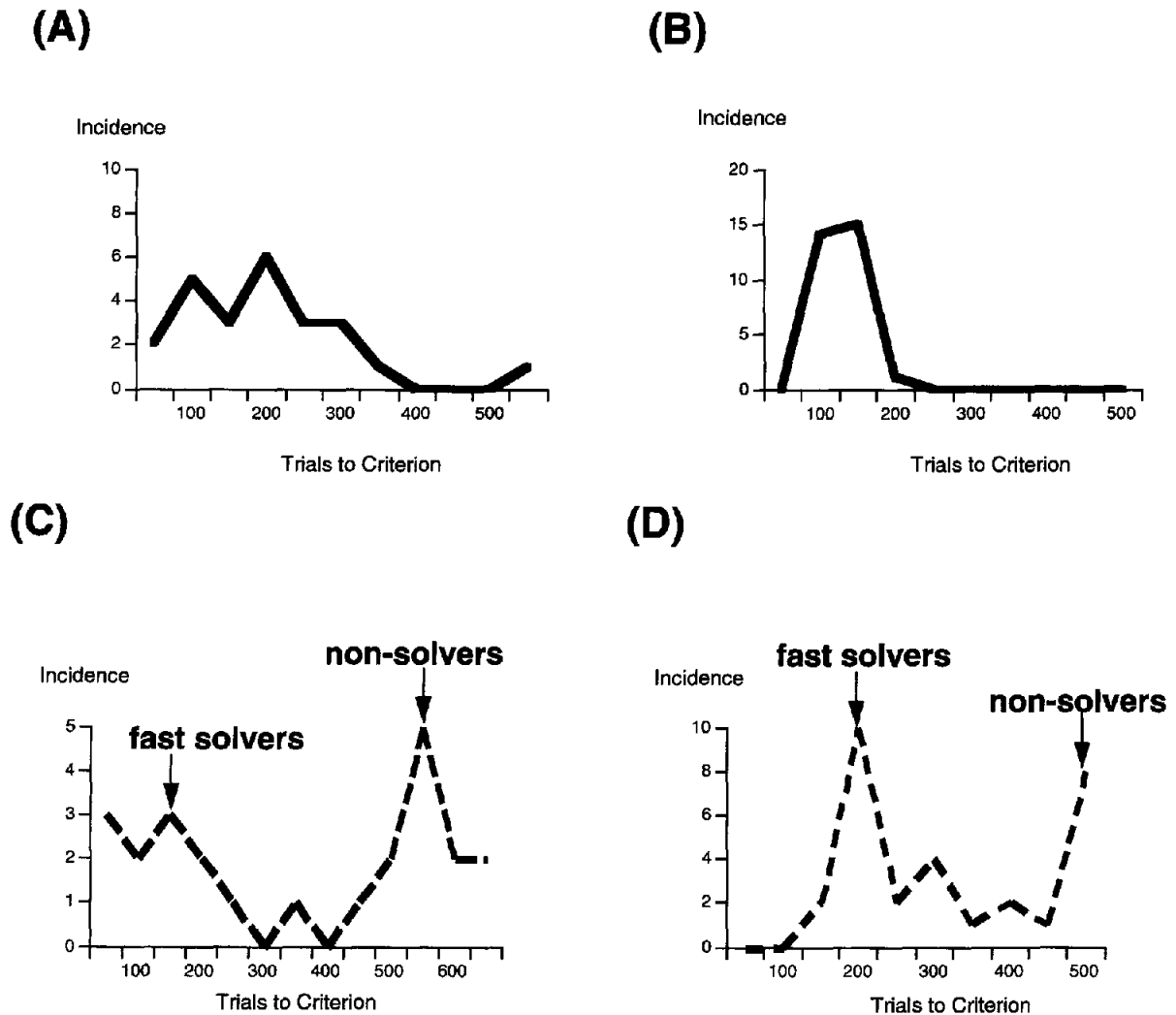


Figure 11. Distribution of learning times on simultaneous discrimination, in terms of trials to reach criterion, pooled over all tasks. A: Control rats show a unimodal distribution, solving most tasks within less than 300 trials (replotted from Eichenbaum et al., 1988). B: The intact model shows a similar distribution of solution times. C: Fornix-lesioned rats show a bimodal distribution, either failing to learn the discrimination or learning just as quickly as control rats (replotted from Eichenbaum et al., 1988). D: The lesioned model shows a similar bimodal distribution, with one peak near 200 and another greater than 500 trials.

that has been trained on the $A+B-$ and $C+D-$ discriminations; for simplicity, only those weights from inputs detailing the spatial locations of A, B, C and D are shown. (These graphs are therefore a subset of the type of information that was averaged to construct Figure 9.) For many of the nodes, many of the weights differ significantly from 0, meaning that those nodes encode information about the spatial locations of these stimuli. For example, consider the weights shown for Node 4. This unit has strong positive weights from inputs active when the stimulus pair AB is present and has strong negative weights activated by stimulus BA. This unit will respond much more strongly to AB than to BA, contributing to a differentiation of stimulus representations for these two trial types. Node 19 performs a roughly complementary function. It is active

when stimulus pair BA is present and is inactive when stimulus AB is present; therefore, it also contributes to differentiating the representations of AB and BA. Note also that these two nodes suffice to allow the discrimination itself to be learned: the model simply has to learn to output a "go-left" response when Node 4 is active and a "go-right" response when Node 19 is active.

Nodes 11 and 16 perform roughly analogous functions for the stimulus pairs CD and DC. Various other nodes shown perform variants on these functions, leading to stimulus representations that are quite different for AB than for BA and for CD than for DC. (Figure 10 is another way of illustrating this claim by showing the differences in internal layer node activation for AB and BA or CD and DC, summed

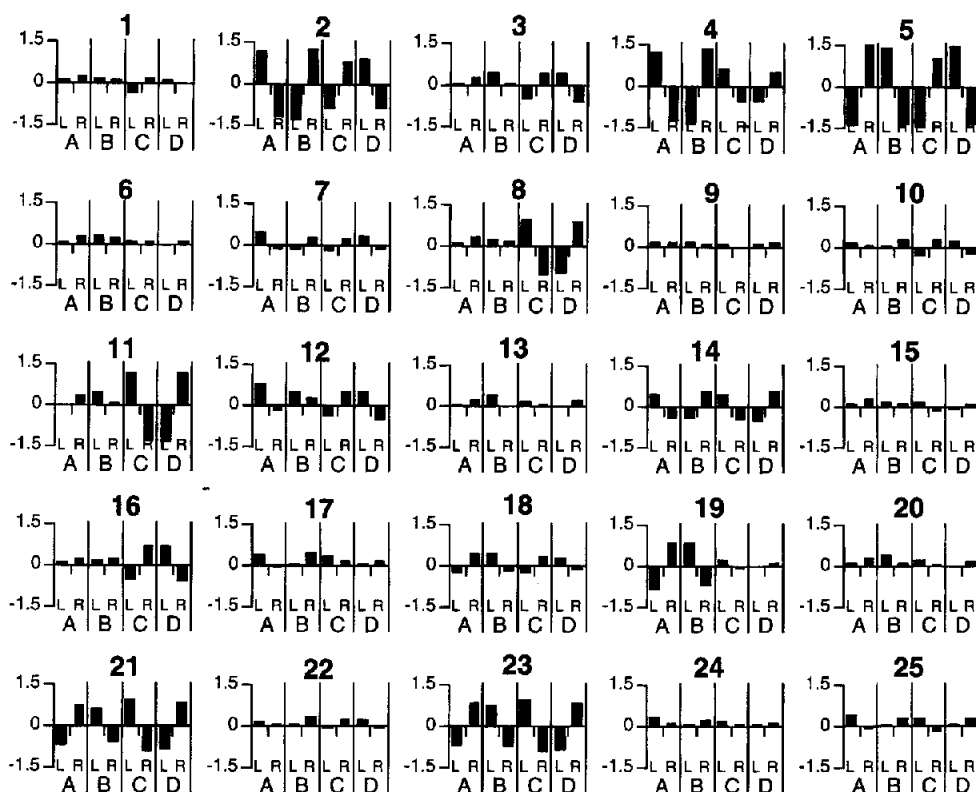


Figure 12. Weights to each of the 25 internal layer nodes in the hippocampal region network of an intact system trained on the A + B- and C + D- discriminations. Only the inputs detailing spatial location (left or right placement) of relevant odors A, B, C and D are shown. Node 4 responds strongly when A is on the left and B is on the right, but not for the opposite spatial placement; Node 19 responds more strongly to the AB placement. Nodes 11 and 16 perform roughly analogous functions for C and D. Some nodes such as 2, 5, and so on, respond to preferred spatial arrangements for both discriminations. The internal layer representation of a particular input pattern is the concatenation of the activation levels of all 25 internal layer nodes. L = left; R = right.

over all the nodes.) These differences increase as training progresses. This increasing difference reflects the construction of internal nodes that are especially sensitive to the spatial locations of these odors. Some nodes (e.g., 2, 5, etc.) have strongly preferred spatial orientations for both odor pairs. Some nodes (e.g., 1, 6, 9, etc.) do not show strong weights to any of the inputs encoding spatial information. This does not imply that these nodes are unused, but merely that they are performing some other function, such as helping the hippocampal region network reconstruct its inputs, which does not make particular use of spatial information.

Interestingly, there is an important correspondence between the activation patterns of the intact model's internal layer nodes and the activation of hippocampal CA1 cells in a rat that is well trained on two odor discriminations. In particular, electrophysiological recordings from the behaving rat reveal some hippocampal cells that respond differentially to a particular locus of response and other cells that respond most strongly when a particular pair of odors is presented in a particular spatial configuration (Wiener et al., 1988). The intact model develops nodes that fall into similar classes, as shown in Figure 12. For example, Node 5 responds strongly when either

positive odor (A or C) is present at the right-hand odor port and thus the correct response is nosepoke right; Node 4 shows the opposite tendency. Node 11 responds most strongly when odor pair CD is present in a particular spatial ordering (C on the left, D on the right); Nodes 16 and 19 respond preferentially to other particular spatial configurations of specific odor pairs. Eichenbaum (1992) concluded that hippocampal cells encode "whatever critical relationships among cues guide accurate performance on the task at hand" (p. 226); this is exactly the result of the representational compression and differentiation mechanisms in the intact model as well.

Now consider a lesioned model that is trained on the same A+B- and C+D- discriminations. It cannot develop representations in its hippocampal region network that differentiate on the basis of spatial information; it can only learn based on the hippocampal network's random recoding of the spatial input and compressed piriform output. As a result, the internal nodes (in the association cortex network) tend not to be able to emphasize spatial information. Figure 13A shows the weights that develop to encode spatial information in the internal layer of a lesioned model that fails to solve either A+B- or C+D-. In strong contrast with the intact weights shown in Figure 12,

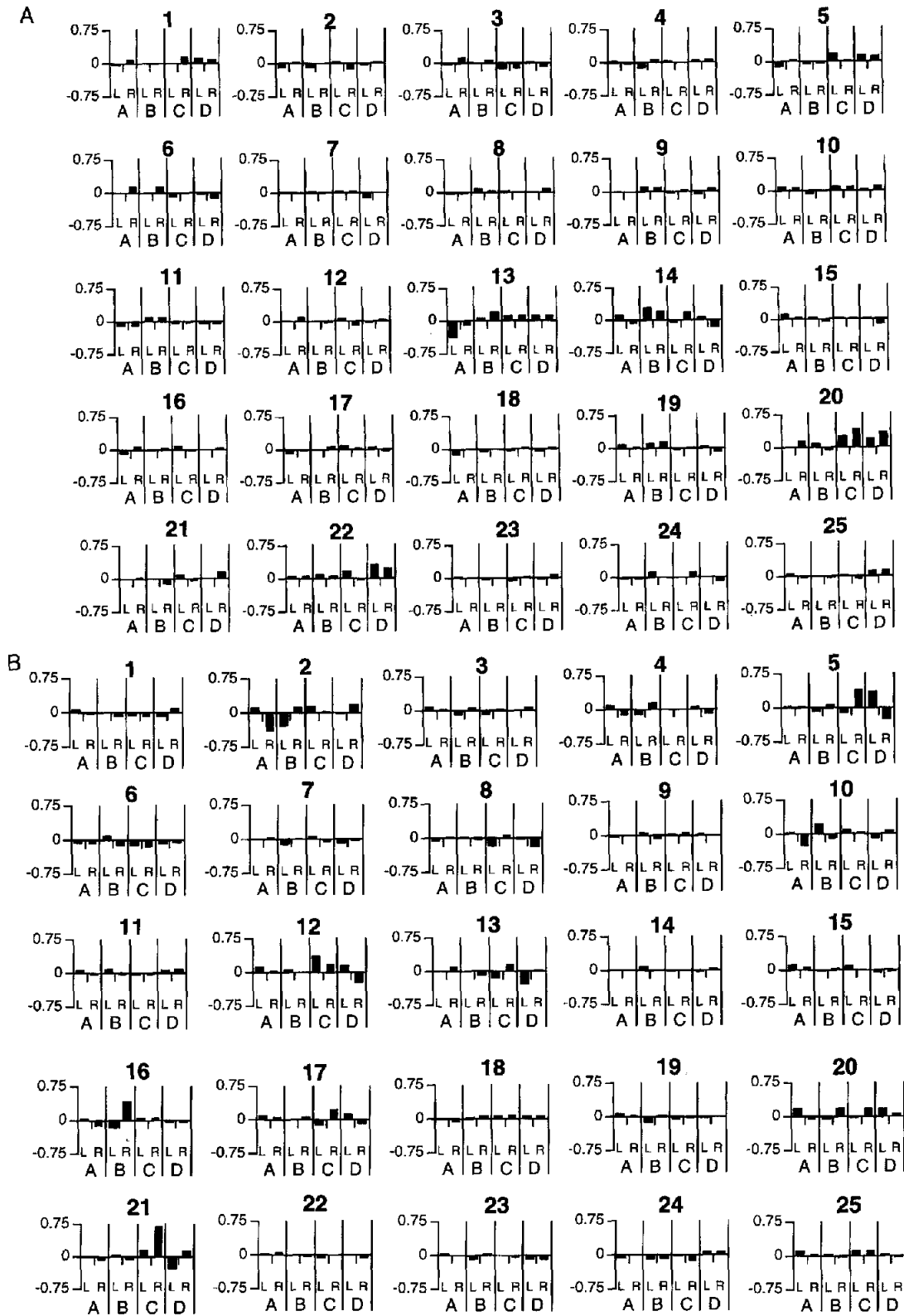


Figure 13. Weights to each internal layer unit in the lesioned network association cortex, from inputs detailing the spatial locations of A, B, C and D, after training on the A + B- and C + D- discriminations. A: In a lesioned system that fails to solve either discrimination, no units develop particularly strong weights from the inputs encoding spatial placement of the relevant stimuli. B: By contrast, in a lesioned system that does solve both discriminations, a few nodes (e.g., 2, 5) do preserve spatial information about the locations of the odors, allowing the discriminations to be solved. L = left; R = right.

no internal layer nodes in the lesioned model develop any particularly strong weights from any of the inputs encoding the spatial locations of the four relevant stimuli. Thus, the exact information necessary to solve the task is not preserved in the internal representation, and the task is not solved.

However, sometimes and at random, the representation does not completely abandon all spatial information. When this occurs, the lesioned network maintains at least some spatial information in its internal representations, and the task may be solvable. Figure 13B shows the weights of the internal layer nodes in a lesioned network that happens to be able to solve $A+B-$ and $C+D-$. This profile is more like the lesioned nonsolver of Figure 13A than like the intact solver of Figure 12: only a few nodes show any selectivity, and, even so, the magnitude of these weights is small. For example, Node 5 shows a slight preference for the DC arrangement over the CD arrangement, and Node 2 shows a slight preference for the BA arrangement over the AB arrangement, but the weights are only a fraction of the strengths of the weights in the intact model. However, they contain enough information to allow the two discriminations to be solved. When either node is activated, the correct behavioral response is "go-right," and otherwise it is "go-left." When this information is present, then, learning the behavioral response is trivial. As a result, learning is quite fast—about as fast as in the intact networks. This explains the pattern of general failure but occasional quick solution on the simultaneous discrimination in the lesioned model that is shown in Figure 11. It should also be noted that on a simple conditioned discrimination, in which stimuli are presented successively and hence no compression is expected, there is no impairment of learning in either the lesioned model (Gluck & Myers, 1993) or lesioned rats (Eichenbaum et al., 1989).

This explanation of why the lesioned model generally fails to solve simultaneous discriminations, but sometimes solves a random discrimination quickly, is similar to the explanation posited by Eichenbaum et al. (1989) to explain the animal data. The fornix-lesioned rats are expected to solve a discrimination only "as a fortuitous consequence of idiosyncratic perceptual variations that permitted each rat to discriminate some left-right odor configurations as distinct stimulus compounds" (Eichenbaum et al., 1989, p. 1214); "for just those problems in which different [spatial] arrangements of the same odors can be perceived as distinct stimulus compounds, the [fornix-lesioned] animals could learn the appropriate go-left/go-right response" (p. 1208).

It should be noted that there is an additional, intertrial difference between the fornix-lesioned and control rats that our trial-level models cannot address: The fornix-lesioned rats show a unimodal distribution of response times, whereas the control rats show a bimodal distribution (Eichenbaum et al., 1989). That is, the control rats appear to sample one of the two cues, decide whether or not to respond, and then sample the second cue and decide whether or not to respond. This gives rise to a bimodal distribution of response times, depending on whether the positive odor is the first or second sampled. In contrast, the fornix-lesioned rats appear to sample both cues together, as if they were determining the correct spatial

response to the compound odor; this gives rise to a unimodal distribution of response times. The current trial-level model cannot address this aspect of the data because it does not currently include any mechanism to simulate response latency. However, these data are consistent with the idea that intact rats perceive stimuli individually, perhaps sampling them sequentially, while fornix-lesioned rats perceive unitary stimulus compounds (Eichenbaum et al., 1989). It is also broadly consistent with the tendency of the lesioned model to overcompress stimuli, while the intact model allows differentiation of stimulus representations.

Probe Mispairings in Rats and Model

Although fornix-lesioned rats can occasionally solve a discrimination as quickly as control rats, this does not necessarily imply that the two populations are using the same mechanisms to learn. Eichenbaum et al. (1989) demonstrated such a difference through an ingenious transfer task: Fornix-lesioned rats that had solved two simultaneous discriminations (e.g., $A+B-$, $C+D-$) were yoked with control rats trained on the same two discriminations. Training then continued on the two solved discriminations concurrently, interleaved with probe mispairings of the familiar stimuli (e.g., $A+D-$, $C+B-$). Control rats continue to perform well on the mispairings, while fornix-lesioned rats perform at chance on the mispairings (see Figure 14A).

The intact and lesioned models can also be applied to this task. For each simulation run, the lesioned model is trained on up to six simultaneous discriminations until it has solved two (e.g., $A+B-$ and $C+D-$). The intact model is then trained on the same two discriminations, thus yoking intact and lesioned model simulations. During a final training phase, the models are presented with probe trials involving the trained stimuli in novel "mispairings" (e.g., $A+D-$ and $C+D-$). Again, the model performance captures many aspects of the animal data.

Like control rats, the intact model shows good performance on the mispairings, as shown in Figure 14B. After reaching criterion on the concurrent discriminations (90% of simulations do this without error), all intact simulations performed perfectly over 10 blocks of training on the mispairings. Like fornix-lesioned rats, the lesioned model shows a drastic impairment. After successfully learning two simultaneous discriminations, the rats have a significantly worse performance on the mispairings than on the trained pairs, although performance is still above chance. By the end of training on the concurrent discrimination, the lesioned model performs at an average 95.4% correct on trained pairings, but then performs at only 84.7% correct on the mispairings, a highly significant difference, $t(9) = -5.85, p < .001$. One discrepancy is that although the lesioned rats do show gradual improvement on the mispairings, the lesioned model does not.

Why does the intact model perform well on probe trials, while the lesioned model does not, even when it can learn the original pairings? The answer again lies in the function computed by individual internal layer nodes. In both the intact model and in a lesioned model that solves $A+B-$ and $C+D-$, hidden nodes evolve that respond preferentially to different

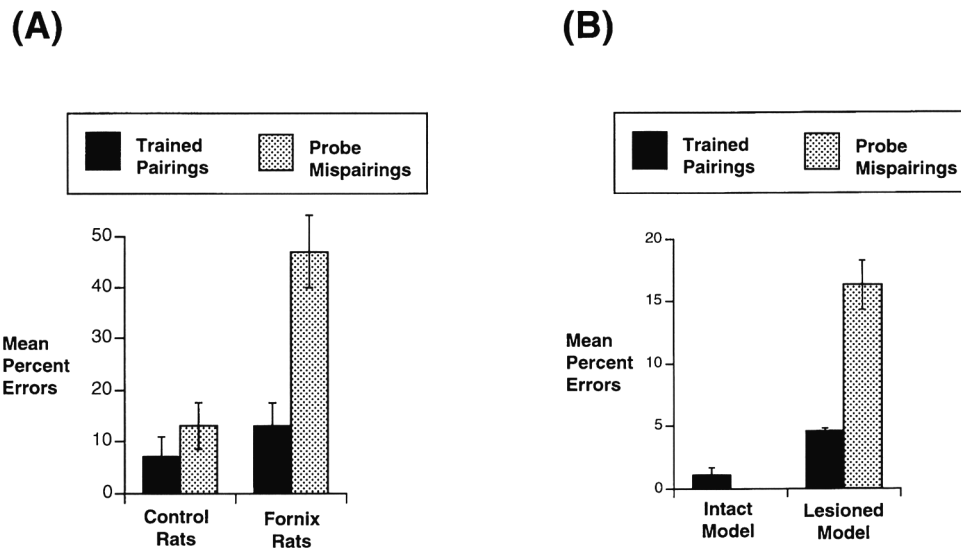


Figure 14. A: Performance on trained pairings and probe mispairings in control and fornix rats, after successful acquisition of two concurrent simultaneous discriminations. Control rats perform as well on mispairings as trained pairings; fornix-lesioned rats perform at chance on mispairings. Replotted from data presented in Eichenbaum et al. (1989). B: Performance on trained pairings and probe mispairings in the intact and lesioned model. The intact model performs well on mispairings, the lesioned model performs much worse, though still well above chance.

spatial arrangements of A and B and of C and D; such nodes appear in Figures 12 and 13B. However, while the lesioned model's nodes tend to respond to spatial preferences regarding either A and B or C and D, the intact model evolves nodes that

have spatial preferences regarding both stimulus pairs. For example, the intact model's internal layer Node 5 (reproduced from Figure 12 in Figure 15A) responds strongly to the BA arrangement but not to the AB arrangement and to the DC

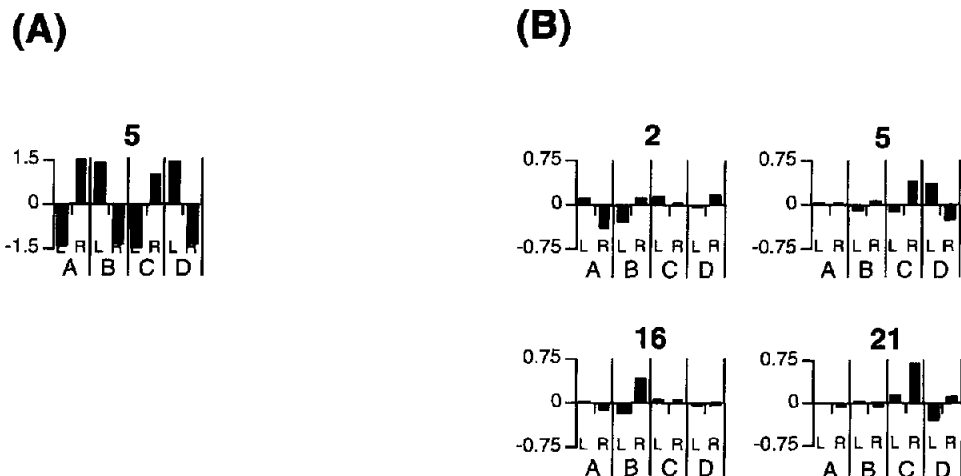


Figure 15. Reproduction of some of the relevant internal layer nodes from Figures 12 and 13B. A: In the intact model hippocampal region network, internal layer Node 5 responds strongly to the spatial arrangement BA and also to the spatial arrangement DC; because this node differentiates both pairs of stimuli, it continues to be strongly activated by stimulus mispairings. B: In the lesioned model association cortex network, internal layer Nodes 2 and 5 respond strongly to particular spatial arrangements of A and B or C and D. Because none of these nodes contains strong weights from more than one pair of odors, none is fully activated by a mispairing. Additionally, some nodes (e.g., 16 and 21) respond most strongly to a single odor in a single location, regardless of what other odor is present. Taken together, the activations of many such nodes may allow the system to respond with at least partial correctness on mispairing trials; this accounts for the better-than-chance performance of the lesioned system on probe mispairings. L = left; R = right.

arrangement but not to the CD arrangement. In effect, this node is active whenever the behavioral response should be "go-right" and inactive whenever the behavioral response should be "go-left." Because this node can differentiate the spatial arrangements of both pairs of stimuli, it will continue to respond to the spatial arrangements of mispaired stimuli. The node will respond strongly to the DA and BC ("go-right") mispairings and not to the AD and CB ("go-left") pairings. Therefore, this node alone carries enough information to allow the intact system to respond correctly on the mispairings, as well as on the trained pairings. To a lesser extent, Nodes 4, 14, and 18 perform similarly.

The few spatially sensitive nodes that evolve in the lesioned system are reproduced (from Figure 13B) in Figure 15B. By contrast with Figure 15A, the lesioned model nodes only contain information about one or the other of the two stimulus pairs. Thus, when a mispairing is presented, none of these nodes can respond fully. For example, when DA is presented, Node 2 will respond partially to the presence of A on the right, and Node 5 will respond partially to the presence of D on the left; but neither node will be fully activated, because Node 2 also looks for the presence of B on the left, and Node 5 also looks for the presence of C on the right. The partial activation of these two nodes may be enough to allow a mild behavioral tendency to "go-right." Additionally, other nodes such as 16 and 21 respond most strongly to the presence of a particular odor in a particular location, regardless of what other odor is present: Node 16 responds strongest to odor B on the left, and Node 21 responds strongest to odor C on the right. These weights are weak, but the partial activations of many such nodes together may also contribute to a correct response. As a result, the lesioned network behavior on the probe trials shown in Figure 14B is above chance; however, transfer is not nearly as good as in the intact model. Further, because these weights are not adapted in the hippocampal region network, performance improves little even with extended training on the mispairs.

Conclusions

We have argued that critical aspects of the simultaneous odor discrimination learning in control and fornix-lesioned rats (Eichenbaum et al., 1988, 1989) can be interpreted via an existing computational model of the role of hippocampal region in associative learning (Gluck & Myers, 1993; Myers & Gluck, 1994). The model assumes that the hippocampal region is able to compress and differentiate stimulus representations according to both stimulus-stimulus and stimulus-outcome regularities. We showed that the model could be generalized to apply to Eichenbaum et al.'s operant paradigm. The intact model, like intact rats, shows facilitation on subsequent discriminations and a unimodal distribution of learning times. A lesioned version of the model, without hippocampal region representational changes (but with piriform-mediated odor compression) captures many aspects of fornix-lesioned rats' performance. The lesioned model correctly shows an overcompression deficit, which impairs discrimination learning; although occasionally, and at random, a particular discrimination may be solvable. When it does solve a task, it does so just

as fast as the intact model. Finally, while both the intact model and control rats show good transfer to novel pairings of trained stimuli, both the lesioned model and fornix-lesioned rats show little transfer. The computational mechanisms and representational processes underlying the model behavior are quite similar to the qualitative explanations proposed by Eichenbaum et al. (1988, 1989) to explain the rat data, namely, the tendency of lesioned rats and model to overcompress or fuse co-occurring odors and to treat them as a single stimulus compound, rather than recognizing distinct cue components. This is an elaboration of the idea presented by Eichenbaum and colleagues (e.g., Eichenbaum et al., 1992) that the hippocampus is necessary for flexible learning that can be expressed in novel situations, such as when familiar odors are presented in novel pairings.

Similarly, Schacter (1985) has noted that human hippocampal-damaged amnesics often seem inflexible or "hyperspecific" in their learning. Although amnesics may be able to acquire new information, they are often unable to express that learning if test conditions differ significantly from learning conditions. This could reflect a kind of overcompression deficit, in which stimulus and context are compressed or fused together, so that the familiar stimulus presented in a new context is unrecognizable and, therefore, the learned association cannot be retrieved. Related ideas have been circulated that the hippocampus is involved in the ability to learn configural tasks in which cue compound and components are assigned different meanings (e.g., Sutherland & Rudy, 1989; Schmajuk & DiCarlo, 1992) and in the ability to utilize contextual cues (Hirsh, 1974; Winocur & Olds, 1978).

Elsewhere, we have proposed that the entorhinal cortex has anatomical and physiological substrates sufficient to perform a particular subfunction of the representational modifications assumed in the intact model (Myers et al., 1995). In particular, we noted the correspondence between superficial entorhinal cortex and superficial layers of the adjacent piriform cortex (Gluck & Granger, 1993), suggesting the possibility of related function. Following the suggestion of Ambros-Ingerson et al. (1990) that the piriform cortex could perform stimulus clustering of olfactory inputs, as described earlier in this article, we suggested that the entorhinal cortex could also perform stimulus clustering based on superficial similarity and also on stimulus co-occurrence (Myers et al., 1995). The remaining aspects of the proposed hippocampal region function, especially predictive differentiation, would then be implemented elsewhere, such as in the dentate gyrus and hippocampus proper. Eichenbaum and Bunsey (1995) have made a similar suggestion that the parahippocampal region (including entorhinal cortex) mediates "fusing" stimuli into compound percepts while the hippocampal formation mediates relational memory processing that allows flexible use of memories. Their concept of entorhinal-mediated fusion includes allowance for temporal fusion across intermediate-term delays.

This hypothesis about the particular contribution of entorhinal cortex to hippocampal region function has implications about the functional effects of various lesion extents. For example, to the extent that a selective lesion of hippocampus spares entorhinal processing, such a lesion might not disrupt hippocampal-dependent tasks that arise from stimulus compres-

sion. One example is latent inhibition: Unreinforced preexposure to a stimulus retards later learning to associate that stimulus with a response (Lubow, 1973). Our intact model shows latent inhibition as a result of stimulus compression. During the preexposure phase, the representation of the stimulus is compressed together with the co-occurring (and equally nonpredictive) background contextual cues. Later, in the learning phase, the stimulus and context must be explicitly redifferentiated to allow selective responding to the stimulus but not to the context alone (Myers & Gluck, 1994). Because this effect is explained in terms of hippocampal-dependent compression processes, it is expected to be eliminated in subjects with hippocampal region damage. Consistent with this prediction, hippocampal ablation abolishes latent inhibition (Kaye & Pearce, 1987; Solomon & Moore, 1975; Schmajuk, Lam, & Christiansen, 1994). By contrast, the hypothesis that the entorhinal cortex is sufficient to underlie stimulus compression argues that a selective hippocampus-only lesion should not eliminate latent inhibition. Consistent with this prediction, latent inhibition survives selective ibotenate hippocampal lesion (Honey & Good, 1993).

How does this argument apply to the simultaneous odor discriminations considered in this article? We have argued that the odor compression in the piriform cortex is sufficient to give rise to overcompression deficits in animals with broad hippocampal lesions. In an animal with a selective hippocampal-only lesion, this behavior should continue, or even be exacerbated by, additional stimulus compression in the entorhinal cortex. Therefore, whether the fornix lesion in the rats tested by Eichenbaum et al. (1988, 1989) is more comparable to a broad hippocampal region lesion or a selective hippocampal-only lesion, the theory still expects overcompression deficits on simultaneous odor discrimination—as long as the piriform cortex survives but the hippocampus is damaged. Consistent with this idea, an entorhinal lesion (which may be assumed to also functionally isolate hippocampus from its major source of sensory input) appears no more disruptive than fornix lesion on discriminations where odors are presented simultaneously (Staubli, Fraser, Kessler, & Lynch, 1986) or successively (Otto et al., 1991).³ There is also some evidence that hippocampal region damage impairs retention and relearning of simultaneous odor discrimination in monkeys (Santibanez & Pinto Hamuy, 1957).

How does this argument apply to other paradigms, especially simultaneous discriminations in other modalities? The answer depends on the extent to which stimulus compression occurs in sensory cortices dealing with other kinds of stimulus. It seems reasonable to assume that many areas of the cortex could process information in similar ways, differing mainly in the particular types of information they process as a result of particular input and output connections (cf. Mountcastle, 1979; Szentagothai, 1975). There is some evidence that stimulus clustering occurs in the auditory cortex (e.g., Bakin & Weinberger, 1990; Weinberger et al., 1990) and in the striate cortex (e.g., von der Malsburg, 1973), which would be consistent with our basic argument. However, the data regarding the effects of hippocampal region damage on nonolfactory simultaneous discriminations are quite mixed. For example, broad hippocampal region lesions can disrupt simultaneous visual

discrimination in monkeys (Mishkin & Pribam, 1954; Pinto Hamuy, Santibanez, Gonzales, & Vicencio, 1957) and dogs (Fuller, Rosvold, & Pribam, 1957). On the other hand, hippocampal ablations in rats do not impair learning to enter arms of a Y maze depending on their brightness (Kimble, 1963) or relearning a simple tactile discrimination (Whishaw & Tomie, 1991). Aside from differences arising from the use of different species, paradigms, and lesion techniques, one possibility for these mixed data is that different sensory modalities are differentially prone to compress the representations of co-occurring stimuli. Some light might be shed on this issue by extending the current model to include additional modules representing the processing in additional sensory cortical areas, and the current work would provide a framework for this kind of extension.

In the meantime, the idea that co-occurring olfactory stimuli may be especially prone to overcompression would be consistent with their chemical nature and high tendency to blend together into compound percepts. Lesioned rats may simply be especially unable to separate out the features of co-occurring olfactory stimuli, resulting in the observed overcompression deficit. The idea that olfactory learning is somehow especially dependent on hippocampal mediation would also be consistent with the fact that the olfactory cortex, alone among sensory areas, has strong direct connections with the hippocampal region (Suzuki, 1994). If this explanation is correct, then there should be much less impairment in hippocampal-lesioned rats on olfactory discriminations where odors do not co-occur. Consistent with this interpretation, fornix-lesioned rats are not impaired on a successive odor discrimination (Eichenbaum et al., 1988) nor on simple discriminations including odors and tactile cues (Whishaw & Tomie, 1991). There might also be less hippocampal lesion impairment on simultaneous odor discrimination if the odors are delivered at sufficient spatial distance from each other; to our knowledge, this remains to be investigated.

Consistent with the rat data, monkeys with hippocampal damage are generally not impaired on nonolfactory simultaneous object discriminations involving brightness, hue, pattern, or objects (Mahut, 1971; Zola-Morgan, Squire & Amaral, 1986). Fornix-lesioned monkeys are also not impaired at acquiring discriminations between pairs of objects (Saunders & Weiskrantz, 1989). Fornix-lesioned monkeys also perform much worse than controls on a transfer task where they are challenged to respond to familiar stimuli in new combinations (Saunders & Weiskrantz, 1989). In this task, monkeys first learn to choose between object pairs (e.g., choose AB+ over AC-); in the transfer task, they are presented with one object (A) and must choose the object that completes a positive pair (e.g., B to form AB+, not C to form AC-). Control monkeys perform well on the transfer task, whereas fornix-lesioned monkeys initially perform near chance. Although the transfer task is complicated, and other issues may be relevant, the

³ The rats in the Staubli et al. (1986) experiment are trained preoperatively on simultaneous discriminations, and so direct comparison with the Eichenbaum et al. (1988, 1989) rats is somewhat hindered. Additionally, the Staubli et al. (1986) task involved choosing maze arms rather than simply nose-poking to odor delivery ports.

result could suggest that, in the original phase, the fornix monkeys tend to overcompress information in the first phase (perceiving AB and AC as compound stimuli) and are unable to recognize and to respond to the components (A, B, and C) in the transfer phase.

However, with broader lesions including entorhinal cortex, monkeys show a transient impairment at simultaneous object discrimination (Zola-Morgan & Squire, 1985; Zola-Morgan, Squire, & Amaral, 1989a). Human hippocampal-damaged amnesics, who also generally have additional extrahippocampal damage, also show an impairment on simultaneous object discrimination (Squire, Zola-Morgan, & Chen, 1988). These data suggest that extrahippocampal structures, such as entorhinal cortex, are critical for learning simultaneous object discriminations.

Unfortunately, in the absence of an experiment like the probe mispairings of Eichenbaum et al. (1989), it is difficult to speculate on whether the impairment in broadly lesioned monkeys and humans results from an overcompression deficit or from some other difficulty. One theoretical possibility is that the entorhinal cortex is required for buffering stimulus information across intermediate-term delays (a few minutes), as suggested by Eichenbaum, Otto, and Cohen (1994). Given this assumption, the deficits in simultaneous discrimination would have more to do with remembering trial-by-trial information than with difficulty distinguishing stimuli. In partial confirmation of this idea, Staubli, Ivy, and Lynch (1984) found that rats with entorhinal lesions were not impaired at simultaneous odor discrimination with 45-s intertrial intervals (ITIs), but the entorhinal-lesioned rats were severely impaired with 3-min ITIs.⁴ An important future extension of our modeling work remains the inclusion of temporal information to explore this idea of entorhinal buffering, as well as addressing other aspects of the simultaneous odor discrimination data, such as the differing response latencies in intact and fornix-lesioned rats.

The aim of this commentary has been to provide a new, computational framework for interpreting existing data on simultaneous odor discrimination in intact and lesioned rats (Eichenbaum et al., 1988, 1989). Not only does the model succeed at accounting for the rat behaviors, but the computational mechanisms involved are broadly consistent with the qualitative interpretations suggested by Eichenbaum et al. to explain their experimental data. We have shown that a model originally conceived to address hippocampal region function in classical conditioning requires only minimal extension to address an operant paradigm, simultaneous odor discrimination. Clearly, there is scope for much more development to allow the model to capture more complex behaviors known to depend on hippocampal region mediation, including temporal

processing, spatial processing, and so on. Nevertheless, even this small step suggests that there may be similar mechanisms underlying hippocampal region involvement in a wide range of paradigms and task demands and that these may eventually be captured within a single, unified computational framework.

References

- Ambros-Ingerson, J., Granger, R., & Lynch, G. (1990). Simulation of paleocortex performs hierarchical clustering. *Science*, *247*, 1344–1348.
- Bakin, J. S., & Weinberger, N. M. (1990). Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. *Brain Research*, *536*, 271–286.
- Berger, T., & Thompson, R. (1978). Neuronal plasticity in the limbic system during classical conditioning of the rabbit nictitating membrane response. I. The hippocampus. *Brain Research*, *145*, 323–346.
- 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.
- Bunsey, M., & Eichenbaum, H. (1993). Critical role of the parahippocampal region for paired association 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.
- Eichenbaum, H. (1992). The hippocampal system and declarative memory in animals. *Journal of Cognitive Neuroscience*, *4*, 217–231.
- Eichenbaum, H., & Bunsey, M. (1995). On the binding of associations in memory: Clues from studies on the role of the hippocampal region in paired associate learning. *Current Directions in Psychological Science*, *4*, 18–23.
- 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 and Neural Biology*, *57*, 2–36.
- Eichenbaum, H., Otto, T., & Cohen, N. (1994). Two functional components of the hippocampal memory system. *Behavioral and Brain Sciences*, *17*, 449–518.
- Fuller, J., Rosvold, H., & Pribam, K. (1957). The effect on affective and cognitive behavior in the dog of lesions of the piriform-amygdala-hippocampal complex. *Journal of Comparative and Physiological Psychology*, *50*, 89–96.
- Gluck, M., & Bower, G. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*, 225–244.
- Gluck, M. A., & 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., & Gochel, J. (1994). A computational perspective on dissociating hippocampal and entorhinal function (Response to Eichenbaum, et al.). *Behavioral and Brain Sciences*, *17*, 478–479.
- Grossberg, S. (1976). Adaptive pattern classification and recoding: Part I. *Biological Cybernetics*, *23*, 121–134.

⁴The ability of Staubli et al.'s (1984) lesioned rats to learn the simultaneous odor discrimination with a short intertrial interval (ITI) does not necessarily contradict Eichenbaum et al.'s (1988, 1989) finding that lesioned rats were severely impaired on simultaneous odor discriminations with an even shorter (10 s) ITI, because Staubli et al.'s lesioned rats had received prelesion training on identical discriminations with different odors, which may have improved their subsequent performance.

- 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.
- Hull, C. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.
- 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.
- Kimble, D. (1963). The effects of bilateral hippocampal lesions in rats. *Journal of Comparative and Physiological Psychology*, *56*, 273–283.
- Kohonen, T. (1984). *Self-organization and associative memory*. New York: Springer-Verlag.
- Levy, W. B., & Steward, O. (1983). Temporal contiguity requirements for long-term potentiation/depression in the hippocampus. *Neuroscience*, *8*, 791–797.
- Lubow, R. (1973). Latent inhibition. *Psychological Bulletin*, *79*, 398–407.
- Luce, R. (1963). Detection and recognition. In R. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103–189). New York: Wiley.
- Mahut, H. (1971). Spatial and object reversal learning in monkeys with partial temporal lobe ablations. *Neuropsychologia*, *9*, 409–424.
- Mishkin, M., & Pribam, K. (1954). Visual discrimination performance following partial ablation of the temporal lobe I: Ventral vs. lateral. *Journal of Comparative and Physiological Psychology*, *47*, 14–20.
- Mountcastle, V. (1979). An organizing principle for cerebral function: The unit module and the distributed system. In F. Schmitt & F. Worden (Eds.), *The neurosciences: Fourth study program* (pp. 21–42). Cambridge, MA: MIT Press.
- Myers, C., Ermita, B., Harris, K., Hasselmo, M., Solomon, P., & Gluck, M. (1996). A computational model of the effects of septohippocampal disruption on classical eyeblink conditioning. *Neurobiology of Learning & Memory*, *66*, 51–66.
- Myers, C., & Gluck, M. (1994). Context, conditioning and hippocampal re-representation. *Behavioral Neuroscience*, *108*, 835–847.
- Myers, C., Gluck, M., & Granger, R. (1995). Dissociation of hippocampal and entorhinal function in associative learning: A computational approach. *Psychobiology*, *23*, 116–138.
- Otto, T., Schottler, F., Staubli, U., Eichenbaum, H., & Lynch, G. (1991). Hippocampus and olfactory discrimination learning: Effects of entorhinal cortex lesions on olfactory learning and memory in a successive-cue, go-no-go task. *Behavioral Neuroscience*, *105*, 111–119.
- Pinto Hamuy, T., Santibanez, G., Gonzales, C., & Vicencio, E. (1957). Changes in behavior and visual discrimination preferences after selective ablations of the temporal cortex. *Journal of Comparative and Physiological Psychology*, *50*, 379–385.
- Port, R., & Patterson, M. (1984). Fimbrial lesions and sensory preconditioning. *Behavioral Neuroscience*, *98*, 584–589.
- Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. Rumelhart & J. McClelland (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.
- Santibanez, G., & Pinto Hamuy, T. (1957). Olfactory discrimination deficits in monkeys with temporal lobe ablations. *Journal of Comparative and Physiological Psychology*, *50*, 472–474.
- Saunders, R., & Weiskrantz, L. (1989). The effects of fornix transection and combined fornix transection, mammillary body lesions and hippocampal ablations on object-pair association memory in the rhesus monkey. *Behavioral and Brain Research*, *35*, 85–94.
- Schacter, D. (1985). Multiple forms of memory in humans and animals. In N. Weinberger, J. McGaugh, & G. Lynch (Eds.), *Memory systems of the brain: Animal and human cognitive processes* (pp. 351–379). New York: Guilford Press.
- Schmajuk, N., & DiCarlo, J. (1992). Stimulus configuration, classical conditioning and hippocampal function. *Psychological Review*, *99*, 268–305.
- Schmajuk, N., Lam, Y.-W., & Christiansen, B. (1994). Latent inhibition of the rat eyeblink response: Effect of hippocampal aspiration lesions. *Physiology and Behavior*, *55*, 597–601.
- Shepard, R. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325–345.
- Solomon, P., & Gottfried, K. (1981). The septohippocampal cholinergic system and classical conditioning of the rabbit's nictitating membrane response. *Journal of Comparative and Physiological Psychology*, *95*, 322–330.
- 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 and Physiological Psychology*, *89*, 1192–1203.
- Squire, L. (1987). *Memory and brain*. New York: Oxford University Press.
- Squire, L., Zola-Morgan, S., & Chen, K. (1988). Human amnesia and animal models of amnesia: Performance of amnesic patients on tests designed for the monkey. *Behavioral Neuroscience*, *102*, 210–221.
- 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.
- Staubli, U., Fraser, D., Kessler, M., & Lynch, G. (1986). Studies on retrograde and anterograde amnesia of olfactory memory after denervation of the hippocampus by entorhinal cortex lesions. *Behavioral and Neural Biology*, *46*, 432–444.
- Staubli, U., Ivy, G., & Lynch, G. (1984). Hippocampal denervation causes rapid forgetting of olfactory information in rats. *Proceedings of the National Academy of Sciences*, *81*, 5885–5887.
- Sutherland, R., & Rudy, J. (1989). Configural association theory: The role of the hippocampal formation in learning, memory and amnesia. *Psychobiology*, *17*, 129–144.
- Sutton, R., & Barto, A. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, *88*, 135–170.
- Suzuki, W. (1994). What can neuroanatomy tell us about the functional components of the hippocampal memory system? (Response to Eichenbaum, et al., target). *Behavioral and Brain Sciences*, *17*, 498.
- Swanson, L. (1979). The hippocampus—New anatomical insights. *Trends in Neurosciences*, *2*, 9–12.
- Szentagothai, J. (1975). The “module-concept” in cerebral cortex architecture. *Brain Research*, *95*, 475–496.

- Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273–286.
- von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14, 85–100.
- Weinberger, N., Ashe, J., Metherate, R., McKenna, T., Diamond, D., & Bakin, J. (1990). Retuning auditory cortex by learning: A preliminary model of receptive field plasticity. *Concepts in Neuroscience*, 1, 91–132.
- Whishaw, I., & Tomie, J.-A. (1991). Acquisition and retention by hippocampal rats of simple, conditional and configural tasks using tactile and olfactory cues: Implications for hippocampal function. *Behavioral Neuroscience*, 105, 787–797.
- Widrow, B., & Hoff, M. (1960). Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention Record*, 4, 96–104.
- Wiener, S., Paul, C., & Eichenbaum, H. (1988). Spatial and behavioral correlates of hippocampal neuronal activity. *Journal of Neuroscience*, 9, 2737–2763.
- Winocur, G., & Olds, J. (1978). Effects of context manipulation on memory and reversal learning in rats with hippocampal lesions. *Journal of Comparative and Physiological Psychology*, 92, 312–321.
- Witter, M. (1993). Organization of the entorhinal-hippocampal system: A review of current anatomical data. *Hippocampus*, 3, 33–44.
- Zola-Morgan, S., & Squire, L. (1985). Medial temporal lesions in monkeys impair memory on a variety of tasks sensitive to human amnesia. *Behavioral Neuroscience*, 99, 22–34.
- Zola-Morgan, S., Squire, L., & Amaral, D. (1986). Human amnesia and the medial temporal region: Enduring memory impairments following a bilateral lesion limited to field CA1 of the hippocampus. *Journal of Neuroscience*, 6, 2950–2967.
- Zola-Morgan, S., Squire, L., & Amaral, D. (1989a). Lesions of the amygdala that spare adjacent cortical regions do not impair memory or exacerbate the impairment following lesions of the hippocampal formation. *Journal of Neuroscience*, 9, 1922–1936.
- Zola-Morgan, S., Squire, L., & Amaral, D. (1989b). Lesions of the hippocampus but not lesions of the fornix or the mammillary nuclei produce long-lasting memory impairment in monkeys. *Journal of Neuroscience*, 9, 898–913.

Appendix

Simulation Details

The simultaneous odor discrimination task is modeled after that used in the empirical studies of Eichenbaum, Mathews, and Cohen (1989). The first phase consists of 500 blocks of training on the A + B– discrimination. Reward is delivered if the output choice matches the spatial location of A on the current trial. One block of training is defined as a set of trials with one presentation of each stimulus compound being trained; in this phase, one block therefore contains one AB trial (A on the left and B on the right) and one BA trial (B on the left and A on the right). Trial order is random within a block. This phase is followed by up to five similar phases of training with new stimulus pairings (i.e., C+D–, E+F–, G+H–, I+J–, K+L–). The rewarded response is the selection of the positive odor of the current pair, regardless of spatial ordering. Learning criterion is defined to be at least 90% correct performance on 10 consecutive trial blocks.

The probe trials are implemented first by training the lesioned system on six discriminations as previously mentioned, and then by selecting two discriminations (e.g., A+B– and C+D–) solved by the system. (No lesioned-model simulation runs failed to reach criterion on a minimum of two discriminations; if simulations solved more than two discriminations, the two were chosen for which criterion was reached in a minimum number of blocks.) These two selected discriminations are then trained concurrently, intermixing trials with each pair in each spatial arrangement, until the 90% correct criterion is again reached. Any simulations that fail to reach criterion on the concurrent task within 500 training blocks are discarded from the analysis. Finally, the mispairings are trained, each block consisting of one trial each of the mispairings (e.g., A+D–, B–C+) in each possible spatial arrangement. The performance reported in Figure 14 compares percent correct over the last 10 blocks of concurrently trained pairs and the first 10 blocks of mispair trials. Intact system simulations runs were then yoked to the lesioned simulations, such that the same probe discriminations were chosen for the *n*th intact simulation as for the *n*th lesioned simulation.

All results described in this article are averaged over 10 different simulation runs.

External Inputs

On each trial, input consisting of a 48-element vector is presented to the system. The first 12 elements represent the 12 possible distinct odor stimuli, A, B, . . . , L. If an odor is present, the corresponding element is activated (set to 1.0); otherwise, it is set to 0.0. The remaining 36 elements comprise 12 three-element subfields, containing location information about each of the 12 possible odors. If a stimulus is present, exactly one element of its subfield is activated, indicating that it is present at the left, center, or right odor port. For the experiments reported here, two stimuli are presented at each trial, one at the left and one at the right odor port (the center odor port is never used in the experiments reported here).

Intact Model

The intact model shown in Figure 6A consists of three interacting modules. The piriform cortex network is assumed to preprocess and cluster odor stimuli according to superficial similarity and co-occurrence. Because the representations of the odors in the original system input are orthogonal, there is no particular basis for clustering based on superficial similarity; however, odors are clustered based on co-occurrence. This piriform network is based on the piriform cortex model described by Ambros-Ingerson, Granger, and Lynch (1990), altered to eliminate feedback input masking that depends on reciprocal connections between piriform cortex and olfactory bulb, not assumed present in our model. The resulting simplified network is similar to the unsupervised, competitive-learning networks proposed by Grossberg (1976), Kohonen (1984), Rumelhart and Zipser (1985), and others.

The piriform network consists of 25 nodes, each receiving 156 inputs: the 12 odor inputs each magnified to occupy a 10-element subfield (all 1s if the odor is present and all 0s otherwise) and the 36 spatial inputs. The connections are then initialized from the uniform distribution $U[0.0 \dots 1.0]$, and then normalized so that the total weight to each node sums to 1.0. The nodes are divided into five patches of five nodes each.

On each trial, each patch individually determines its winning node j with greatest activation y_j determined as:

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

where w_{ij} is the connection strength from input i to unit j , y_i is the activation of input i , and $f(x) = 1/(1 + e^{-x})$. The output of each winner j is set to 1.0, and the outputs of all other nodes in the patch are set to 0.0. Each node j then updates its weights as:

$$\Delta w_{ij} = \beta(t_j - y_j)y_i,$$

with learning rate $\beta = 0.005$, $t_j = 1.0$ for winning nodes and $t_j = 0.0$ for all other nodes, and all weights are bounded as $0.0 \leq w_{ij} \leq 1.0$.

The second module, a hippocampal region network, receives as input the output from the piriform network as well as the external inputs detailing spatial information. This is a fully connected predictive autoencoder with 61 inputs (36 spatial external inputs plus 25 outputs from the piriform network), 25 internal layer nodes, and 63 output nodes trained to reconstruct the 61 inputs plus a prediction of whether the behavioral response was "go-left" or "go-right." The activation of each internal layer and output node j is computed as:

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

where w_{ij} is the connection strength to j from a node i in the previous layer, θ_j is the bias of node j , and $f(x) = 1/(1 + e^{-x})$. The w_{ij} and θ_j are initialized from a uniform distribution $U[-0.1 \dots +0.1]$; two random weights on each internal layer node are initialized as $U[-1.0 \dots +1.0]$. The training signals t_i to the first 61 output units are simply the 61 inputs, while the training signals t_i to the last 2 output units were 1.0 for the chosen output response and 0.0 for the unchosen output. The weights w_{ij} are trained by error backpropagation (Rumelhart, Hinton, & Williams, 1986), such that on each trial:

$$\Delta w_{ij} = \beta \text{err}_j y_i (1 - y_j) + \alpha \Delta' w_{ij},$$

with learning rate $\beta = 0.25$ and momentum $\alpha = 0.9$; biases θ_j are trained as if they are weights from a unit i that constantly outputs $y_i = 1.0$; and $\Delta' w_{ij}$ represents the last change to weight w_{ij} . For output units, $\text{err}_j = t_j - y_j$, whereas for internal layer units,

$$\text{err}_j = \sum_k \text{err}_k y_k.$$

The external output of this module is a 25-element vector, specifying the activations of the hippocampal network internal layer nodes, and represents the new representation formed by the hippocampal region network.

The association cortex network contains 61 input nodes, which receive the same inputs as the hippocampal region network: 36 spatial external inputs plus the 25 outputs of the piriform network. It also contains 25 internal layer nodes and 2 output nodes L and R corresponding to the possible left and right output response choices. Initialization and activation functions are identical to the hippocampal region network. The actual system behavioral response, a choice of the left or right odor port, is computed from the 2 output nodes L and R according to: $Pr(\text{response} = \text{choose left}) = 1/1 + e^{\phi(y_R - y_L)}$ with $\phi = 10.0$. The output units are trained as in the hippocampal region network, except that $\beta = 0.5$ and there is no momentum ($\alpha = 0.0$). The internal units are trained similarly except that the error err_j on an internal layer node j in the association cortex network is computed as the difference between y_j and the activation of the j th hippocampal region network internal layer node. In this model, for simplicity, there is a one-to-one mapping between the internal layer nodes of the hippocampal region and association cortex networks; elsewhere we have shown that this one-to-one mapping is not necessary and a fixed linear recombination suffices (Gluck & Myers, 1993).

The number of nodes in each of the model's networks is essentially arbitrary. After the size of the input and output vectors was fixed, the remaining parameters of a number of units in the hippocampal network hidden layer, association cortex hidden layer, and piriform network were set equal for simplicity. Although we did not conduct parametric studies regarding the optimal value for this number, informal tests showed that the model performance was not particularly fragile with respect to the number of nodes. The most critical parameters are the width of the patches in the piriform network, because this governs the amount of compression that will occur across inputs, and the variance in weight initialization, because a lesioned system with too little variance will never solve any discriminations, whereas too much variance will allow solution of all discriminations. The parameters used here represent a trade-off between these tendencies.

Lesioned Model

The lesioned model assumes that the hippocampal region network is disrupted to preclude weight adaptation. This is implemented by reducing the hippocampal learning rate to 0.0. As a result, the internal representations in the hippocampal region network are a random (fixed) transformation of the piriform cortex output and the spatial input. This is a significant departure from the simpler cortico-hippocampal model presented in Gluck & Myers (1993) that implemented the lesioned model by assuming the absence of all hippocampal region stimulus preprocessing and assuming that cortical network internal layer representations were fixed.

Received June 9, 1995

Revision received August 14, 1995

Accepted December 20, 1995 ■