9 Entorhinal Cortex

Chapter 8 argued that the anatomy and physiology of cerebral cortex was consistent with self-organization, including the development of topographic maps and clustered stimulus representations. Because so much of cortex shares a uniform stylized structure, it seems likely that other areas of cortex could perform similar functions. Various cortical regions differ chiefly in the kind of information they process, depending on their inputs and outputs.[1] Thus, primary auditory cortex processes sound because it receives sound inputs, while primary motor cortex guides movements because its outputs connect to motor neurons in the body. Although this scenario is doubtless oversimplified, it still provides a useful way for thinking about—and modeling—cortical function.

The hippocampal region as defined in this book includes the entorhinal cortex. Some definitions of the hippocampal region also include the nearby perirhinal and parahippocampal cortices. Each of these cortical areas may well perform information-processing functions similar to those described in chapter 8. The difference between, say, entorhinal and piriform cortex would be that, whereas piriform cortex receives and operates on odor inputs, the entorhinal cortex receives highly processed inputs covering the full range of sensory modalities. On the assumption that cortex is capable of self-organizing and clustering inputs, this would suggest that entorhinal cortex should be able to compress information across sensory modalities—for instance, combining the visual, auditory, olfactory, and other features of a stimulus.

This chapter considers three computational models that propose possible functions for entorhinal cortex and its interaction with other hippocampal-region and cortical structures. The first two concentrate on the entorhinal cortex as an input preprocessor for the hippocampus; the third is concerned with the storage of hippocampal-derived representations (or "memories") into cortical long-term storage sites. Finally, the chapter discusses a qualitative theory that focuses on the entorhinal cortex and that is consonant with some of the computational modeling.

9.1 ANATOMY AND PHYSIOLOGY OF THE HIPPOCAMPAL REGION

It sometimes seems as if every scientist studying the hippocampus has a different definition for specifying which structures constitute the hippocampal region (or hippocampal formation, hippocampal system, etc.). In this book, we have defined the **hippocampal region** to include the hippocampus (including subfields CA1 and CA3), dentate gyrus, subiculum, and entorhinal cortex (and possibly fimbria/fornix).* Other definitions would be equally valid for different anatomical, physiological, and behavioral purposes. Our definition of the hippocampal region as a functional unit is useful for two reasons. First, all the included structures appear to contribute to memory functions that are independent of sensory modality; thus, they appear to constitute a multimodal memory system. Second, until very recently, many techniques for "hippocampal lesion" actually involved damage to the broader hippocampal region. Hence, many early data on the effects of "hippocampal lesion" more accurately describe the effects of broader and more inclusive "hippocampal-region damage."

The basic routes of sensory information flow within the hippocampal region are schematized in figure 9.1. Sensory input, such as smell or auditory information, is processed first by primary sensory cortex and other cortical areas devoted to that modality (called **unimodal sensory cortices**) before progressing to **polymodal association cortices,** which integrate information across multiple sensory modalities. From there, information enters the entorhinal cortex. The one major exception to this rule, as we noted earlier, is that odor information from the piriform cortex travels directly to the entorhinal cortex. The entorhinal cortex also receives input from subcortical areas such as the septum, thalamus, hypothalamus, and amygdala.

Entorhinal cortex is intermediate in structure between six-layered neocortex and two-layered allocortex. For this reason, it is sometimes classified with the rather unwieldy name of **periallocortex.** For current purposes, it is sufficient to note that the cell layers in the entorhinal cortex are rather indistinct and clustered into **superficial layers** (layers II and III) and **deep layers** (layers IV–VI). Pyramidal neurons are located in both the superficial and deep layers. Sensory input projects largely to the superficial layers.

From superficial entorhinal cortex, axons project to the dentate gyrus and hippocampus (figure 9.2). The tract from entorhinal cortex to dentate gyrus is

*Our definition is roughly consistent with that of two premier hippocampal anatomists, David Amaral and Menlo Witter (1989), although they group the same structures under the term *hippocampal formation*.

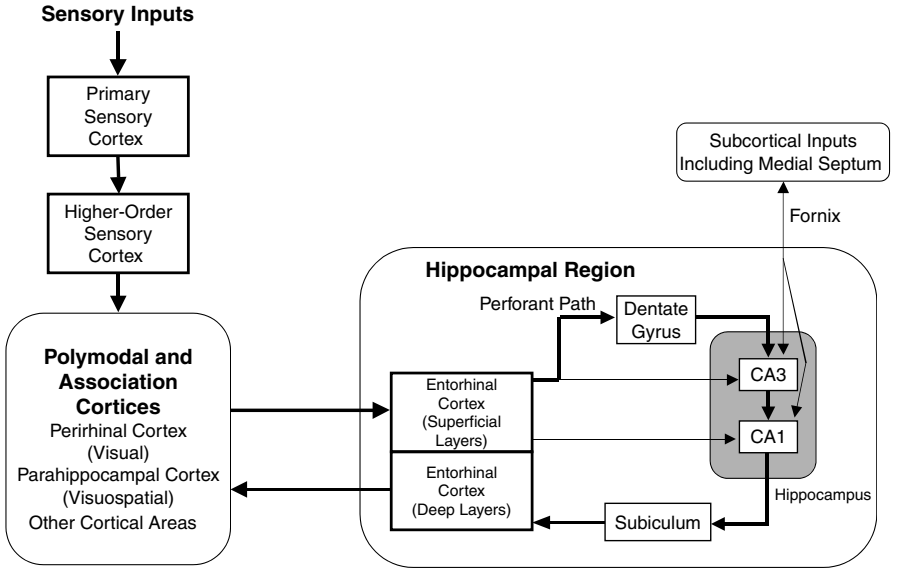


Figure 9.1 Sensory input travels to primary sensory cortices, then to higher-order association cortices, eventually reaching entorhinal cortex. There is also a direct path from primary olfactory (piriform) cortex to entorhinal cortex. Superficial layers of entorhinal cortex project to the dentate gyrus via the perforant path and also directly to hippocampal subfields CA3 and CA1. Information flows in a largely unidirectional fashion from dentate gyrus to CA3 to CA1, through subiculum, and back to the deep layers of entorhinal cortex. From there, information projects back to the cortical areas where it originated. There is also an input/output pathway through the fornix that connects hippocampus with subcortical areas including the septum, thalamus, hypothalamus, and amygdala.

known as the **perforant path,** because these fibers must travel through (perforate) the subiculum to reach their destination. From dentate gyrus, axons travel to hippocampal subfield CA3, which also receives a direct projection from entorhinal cortex. Recall from chapter 5 that the combination of strong dentate and weaker entorhinal projections to CA3, together with high recurrency among CA3 neurons, suggests that CA3 might function as an autoassociator.

Information next projects from CA3 to hippocampal field CA1, through the subiculum, back through the deep layers of entorhinal cortex, and finally back to the same cortical areas that gave rise to the information in the first place. There are additional pathways other than those described here or shown in figure 9.1, but this unidirectional flow from entorhinal cortex through dentate gyrus, hippocampus, and subiculum back to entorhinal cortex is a primary feature of the circuitry.[2]

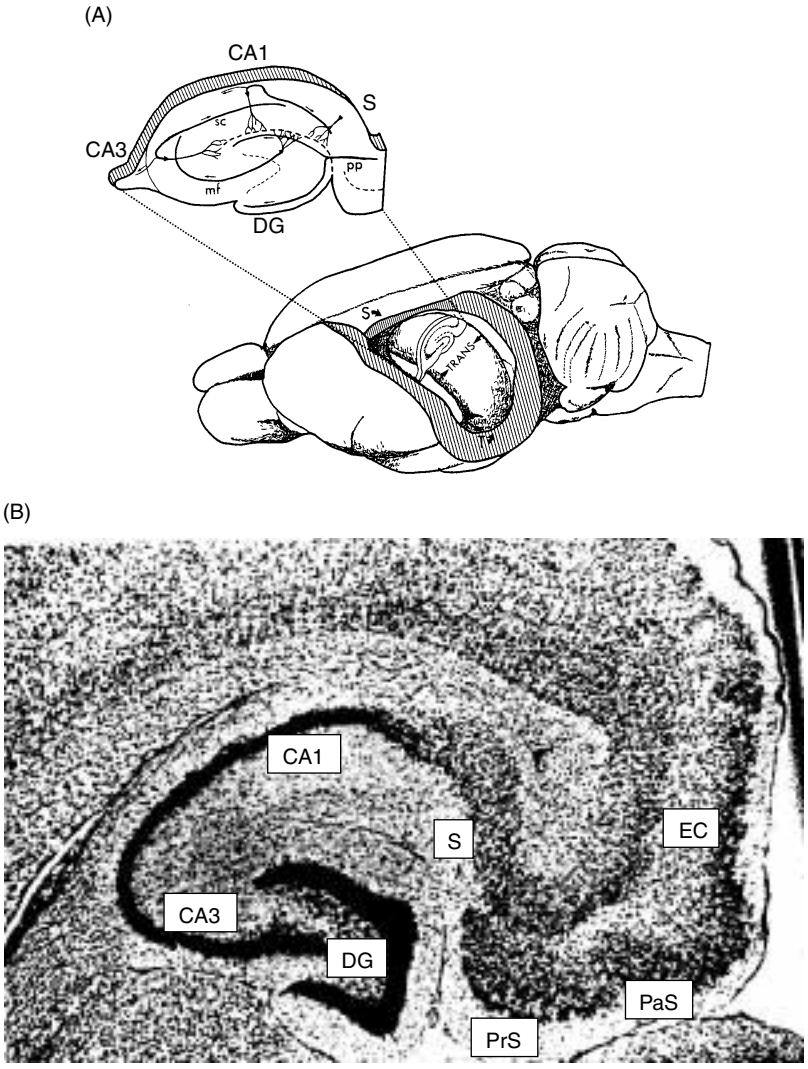


Figure 9.2 (A) Drawing of the rat brain, with cortical surface removed to show position of the hippocampus. A transverse slice through the hippocampus reveals the "C"-shaped organization (cutaway). Hippocampal fields CA1 and CA3, dentate gyrus (DG), and subiculum (S) are as shown. A few information pathways are drawn to illustrate the roughly unidirectional flow (pp = perforant path, mf = mossy fiber, sc = Schaeffer collaterals). (Reprinted from Amaral & Witter, 1989, figure 2.) (B) Photomicrograph of a section through the rat hippocampal region. The hippocampus (including fields CA1 and CA3) is visible as a "C"-shaped line of principal cells; the dentate gyrus (DG) is a backward "C"-shaped line of principal cells. Adjacent to hippocampal field CA1 is the subicular complex, including subiculum (S), presubiculum (PrS), and parasubiculum (PaS). The entorhinal cortex (EC) is the primary pathway for sensory information traveling into and out of the hippocampal region. (Adapted from Amaral & Witter, 1989, figure 1.)

For many years, it was believed that the hippocampal region functioned as a processing chain and that damage to any one link in the chain would disrupt processing in the entire chain. Further, there appeared to be a law of mass action, meaning that impairment increased in an orderly fashion as lesion size increased. Figure 9.3 shows one example of this apparent rule: On a battery of four measures of memory, there was progressively worse performance in monkeys with increasing hippocampal-region damage.[3] However, the story is not quite so simple; later studies have shown that there may be qualitative, not just quantitative, effects of different lesion extents.

An important methodological advance in recent years has been the development of **neurotoxic** lesions, such as the injection of **ibotenic acid,** which destroys neuron cell bodies near the injection site but spares nearby cell bodies as well as any axons that may be passing through the area.[4] This is in contrast to older lesion techniques, particularly surgical removal of tissue by **ablation** or **aspiration,** which not only removed all cells and fibers in the

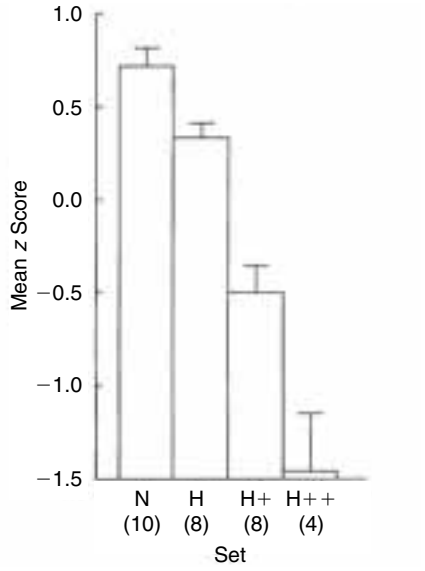


Figure 9.3 On many tasks, the severity of memory impairment is proportional to the extent of hippocampal-region lesion. On a battery of four measures, including delayed nonmatch to sample (DNMS) and delayed retention, monkeys show worse performance with increasing lesions. (Higher mean *z*-score indicates better performance over all four measures.) N = normal (ten monkeys); H = lesion of hippocampus, dentate gyrus, and subiculum (eight monkeys); H+ = H lesion plus adjacent entorhinal and parahippocampal cortices (eight monkeys); H++ = H+ lesion plus anterior entorhinal cortex and perirhinal cortex (four monkeys). (Reprinted from Zola-Morgan, Squire, & Ramus, 1994, figure 4.)

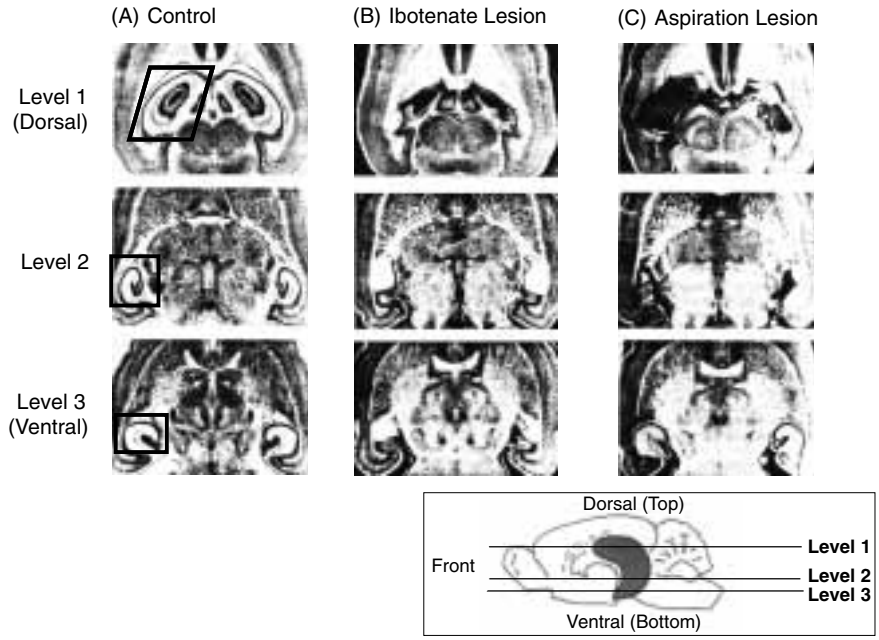


Figure 9.4 Photomicrographs of the rat brain, taken in a horizontal plane, at three levels ranging from dorsal (level 1) to ventral (level 3); the top of each picture is the front of the brain. Inset shows a schematic of a rat brain, with hippocampus in dark gray, illustrating the position of levels shown in A–C. (A) A control rat. Cell bodies in the hippocampus are visible at all three levels (left hippocampus is outlined in black at each level). (B) Hippocampal lesion by injection of ibotenic acid to destroy cell bodies. The lesion is visible in the picture as white spaces where the hippocampal cell bodies should be. Destruction of the hippocampus is relatively complete at all levels, and there is little damage to adjoining regions. (C) Aspiration lesion of the hippocampus. Areas where tissue has been removed are shown as black. Compared with the ibotenate lesion in (B), the aspiration lesion results in variable degrees of damage to the hippocampus; thus, while the dorsal hippocampus (top row) is largely destroyed, the ventral hippocampus (bottom row) is largely spared. Conversely, the aspiration lesion is apt to destroy nearby tissue; thus, in the top row, the black area extends beyond the hippocampus and into nearby tissue. (A–C adapted from Jarrard & Davidson, 1991, figure 4; inset adapted from West, 1990, p. 6, figure 1.)

target area, but also often involved cutting out significant portions of tissue overlying the target area.

Figure 9.4 compares ibotenic acid and aspiration lesions of hippocampus. The ibotenic acid lesion, shown in figure 9.4B, results in relatively complete destruction of cell bodies in the hippocampus and dentate gyrus (seen as white patches devoid of darkly stained cell bodies). In contrast, the aspiration lesion—visible as "empty" dark patches in figure 9.4C—results in a less

accurate removal of tissue. In the example shown, the aspiration lesion generates nearly complete damage to the left dorsal (upper) hippocampus but incomplete damage to the left ventral (lower) hippocampus and much less damage to the right hippocampus. Additionally, the aspiration lesion extends somewhat beyond the hippocampus into nearby brain areas.

Thus, the behavior of an animal with aspiration lesions of hippocampus may be misleading for at least two opposing reasons: The residual hippocampal tissue may allow the animal to maintain some hippocampal-dependent processing, while ancillary damage to other nearby structures may result in deficits that are not hippocampal-related. Neurotoxic lesions, particularly ibotenic acid lesions, minimize these problems, allowing a more accurate analysis of how an animal behaves with specific and localized hippocampal damage. More recently, ibotenic acid has also been used to create specific and precisely localized entorhinal lesions.

By comparing memory deficits in animals with specific hippocampal (or entorhinal) lesions against animals with larger hippocampal-region lesions, researchers are beginning to understand how these separate structures interact in normal functioning. At the same time, these ideas are being explored via computational models. In the next section, we turn to review some models of entorhinal and hippocampal interaction in learning and memory.

9.2 COMPUTATIONAL MODELS

Only recently have computational models begun to address the individual functional roles of different hippocampal-region structures and how these structures might contribute to a coherent learning and memory system. As we discussed in chapter 5, hippocampal field CA3 has received extensive study owing, in part, to its physical resemblance to an autoassociative network. However, there has been much less consideration of how CA3 might interact with other hippocampal-region structures and with the rest of the brain. With a few notable exceptions, there has been even less study of the other hippocampal-region structures.

More recently, the entorhinal cortex has begun to receive attention from computational modelers for two reasons: first, because of its similarity to neocortex, which has already been extensively modeled, and second, because of exciting new empirical data, which suggest that many behaviors that were previously thought to depend on the hippocampus may actually be mediated by the entorhinal cortex.

In this section, we review three computational models of entorhinal cortex. The first is an elaboration of our own cortico-hippocampal model, which

presumes that the entorhinal cortex plays a specific role in compression of redundant stimulus representations. The second is an extension of the Schmajuk-DiCarlo (S-D) model, which assumes that the entorhinal cortex is critical for stimulus configuration. Finally, a model proposed by Edmund Rolls suggests how information processed in the hippocampal region might be transferred from entorhinal cortex back to the neocortex.

Entorhinal Cortex and Redundancy Compression

In the rat, the piriform and entorhinal cortices lie adjacent to each other and are so similar anatomically that it is difficult to determine precisely where one structure ends and the other begins. The superficial layers of the piriform and entorhinal cortices are especially similar in anatomy and physiology, leading many researchers to suggest that the two structures may share a related function.[5] The most obvious difference between these two cortical areas is in their inputs: As was noted above, piriform cortex receives primarily olfactory input, whereas entorhinal cortex receives highly processed information from the full range of sensory modalities as well as from multimodal association areas (figure 9.1). In a sense, the entorhinal cortex represents the highest stage of cortical processing, in which all sensory information converges. Entorhinal cortex would thus be a logical place to perform clustering among stimuli in different modalities (e.g., smell and vision) or among the multimodal features of a single stimulus.

We have proposed that the entorhinal cortex can be modeled as follows:[6] The superficial layers of entorhinal cortex receive highly processed, multimodal stimulus inputs. Entorhinal cells are grouped together into clusters; within each cluster, cells compete to respond to the input (figure 9.5A). Cells that win this competition undergo plasticity, making them more likely to respond to similar inputs in the future (figure 9.5B). Meanwhile, losing cells also undergo plasticity, making them less likely to respond to similar inputs. These cells then become more likely to win the competition in response to very different inputs (figure 9.5C).

Our entorhinal network model performs unsupervised clustering of its inputs, just as the original piriform model did. Similar inputs evoke similar responses. Additionally, the network performs pattern compression or clustering. For example, suppose the pattern trained in figure 9.5C is really a compound stimulus consisting of three component stimuli that always cooccur. The network learns a response to the compound without distinguishing the components. Later, if one of the components is presented alone (figure 9.5D), it will tend to activate the same nodes as did the compound. Moreover, it may activate these nodes less strongly than did the compound.

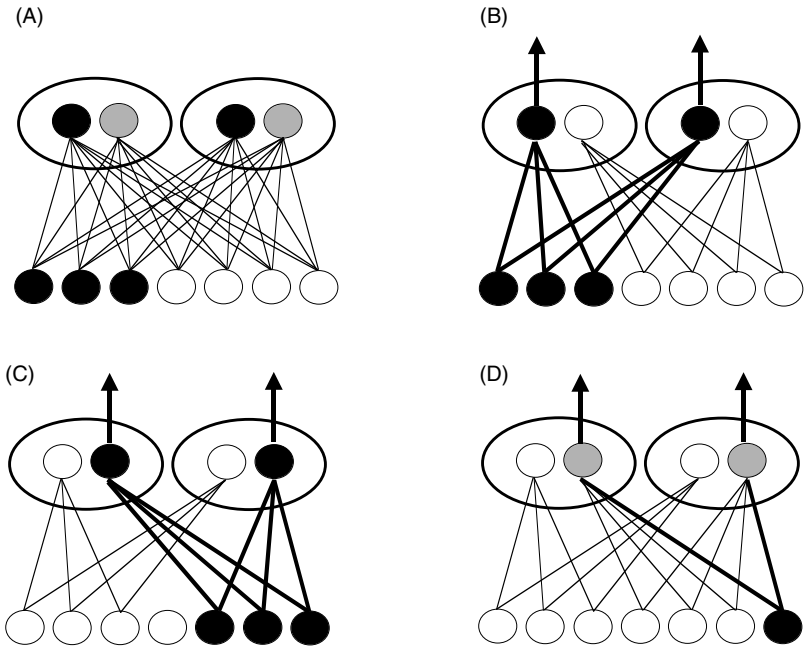


Figure 9.5 (A) Schematic of entorhinal cortex model. A layer of nodes is organized into clusters (two clusters containing two nodes each are shown; the full model consists of 100 nodes organized into five 20-node clusters). Within a cluster, nodes compete to respond to input patterns. If an input pattern is presented (the three black input nodes represent stimulus features that are present on the current trial), one node within each cluster responds most strongly: It wins the competition. (B) Winning nodes are allowed to output, while inhibition silences other nodes in the clusters. The winning nodes undergo plasticity, making them more likely to respond to similar inputs in the future, while losing nodes undergo plasticity to make them less likely to respond to similar inputs in future—and hence more likely to respond to different kinds of inputs. (C) When a very different input pattern is presented, other nodes are likely to win the competition. (D) The network has no way of knowing whether the trained pattern represents one complex stimulus with many different features or a compound of several individual stimuli. Suppose the pattern trained in (C) is really a collection of three different component stimuli. Now, if one component stimulus is presented, this will tend to maximally activate the same nodes in each cluster as the compound stimulus did. The activation may be weaker than that to the pattern trained in (C), but as long as it is stronger than the activation of other nodes in the cluster, it will be sufficient to win the competition. Hence, the network tends to generalize from compound stimuli to the components. This means that the network compresses together the representations of co-occurring stimuli (and stimulus features) into complex compound stimuli.

However, as long as these nodes are more strongly activated than other nodes in the cluster, they will continue to win the competition. Thus, the network tends to represent the components in the same way as it represented the compound. In effect, the network performs pattern completion; when

presented with a component in isolation, it tends to evoke the representation of the full compound stimulus on which it was trained.

A Model of Selective Hippocampal Lesion That Spares Entorhinal Cortex. The implication of this network model is that the entorhinal cortex alone should be sufficient to implement one aspect of the hippocampal-region function proposed in our cortico-hippocampal model from chapter 6: representational compression of co-occurring (redundant) stimuli.[7] Other proposed hippocampal-region functions (particularly representational differentiation) would be mediated elsewhere in the hippocampal region; we will return to this issue later in this chapter.

For now, we consider how an entorhinal network might interact with cortical and cerebellar learning systems in the absence of other hippocampal-region processing. Figure 9.6 shows a schematic of information flow in the hippocampal region, simplified from figure 9.1. So far, we have focused mainly on the behavioral effects of a broad lesion of the entire hippocampal region extending from the entorhinal cortex through to the hippocampus (the **HR lesion;** figure 9.6B). But the entorhinal cortex has input and output connections with cortical areas that are independent of its connections with hippocampus. This implies that a selective lesion of the hippocampus and possibly dentate gyrus and subiculum (the **H lesion;** figure 9.6C) may spare some or all entorhinal processing. Note that the converse is not true: If the entorhinal cortex is lesioned (the **EC lesion;** figure 9.6D), the hippocampus may be functionally isolated from sensory input, and so the effects may be just as dramatic as a full HR lesion.

Recall that our intact cortico-hippocampal model of figure 9.7A assumes that all hippocampal-region representational processing takes place within a hippocampal-region network. This hippocampal-region network performs redundancy compression and predictive differentiation, creating new stimulus representations that are adopted by the cortico/cerebellar network. The HR-lesion model of figure 9.7B—which assumes that all of the hippocampal-region mediated processing is disabled—is really an "H + EC lesion" model. We can now consider an intermediate system: an H-lesion model, in which the hippocampus is lesioned but the entorhinal cortex is spared (figure 9.7C). Note that since the dentate gyrus, hippocampus, and subiculum form a largely unidirectional processing chain (refer to figure 9.1), damage to the hippocampus is liable to disrupt the dentate gyrus (by disabling its primary output pathway) and the subiculum (by disabling a primary input pathway). Hence, in the H-lesion model, the entorhinal cortex may be the only hippocampal-region structure that is still functional.

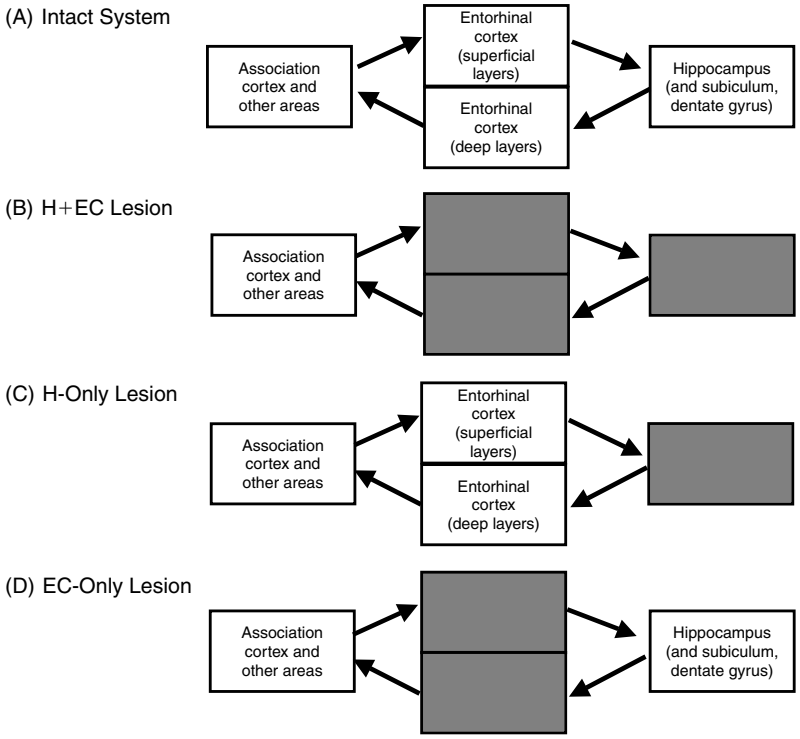
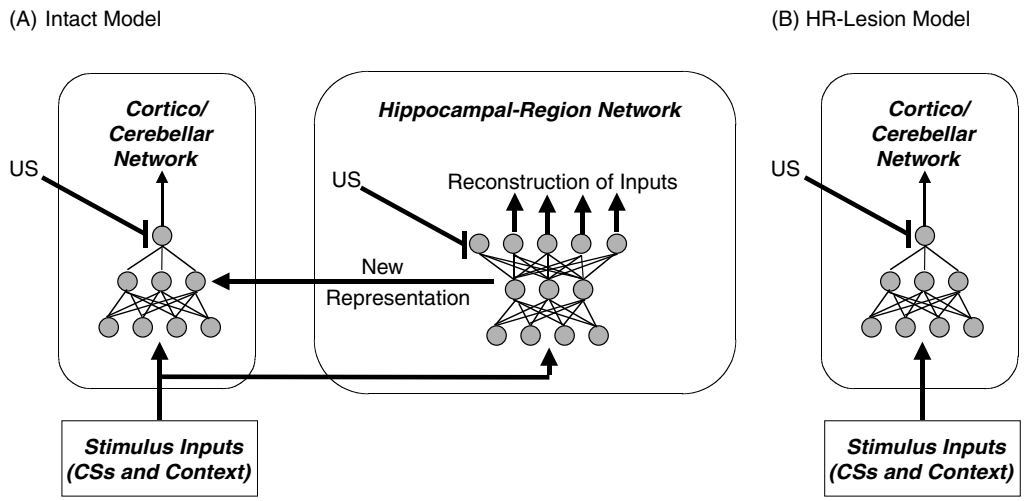


Figure 9.6 (A) In the normal, intact brain, information flows from cortical association areas and other cortical areas to entorhinal cortex to the hippocampus and back out through the entorhinal cortex. (B) Broad hippocampal-region damage (the H+EC lesion) is assumed to destroy processing in entorhinal cortex, hippocampus, and other hippocampal-region structures. (C) A more selective lesion of the hippocampus (H-only lesion) that leaves the entorhinal cortex intact may not totally disrupt processing in the entorhinal cortex; (D) however, a selective lesion of entorhinal cortex (EC-only lesion) functionally disconnects the hippocampus from its sensory input/output pathway. The behavioral effects of EC-only lesion may thus be comparable to those of a full H+EC lesion.

Therefore, in our H-lesion model of figure 9.7C, we assume that the entorhinal network performs stimulus compression and provides these compressed representations to the cortico/cerebellar network. Other representational changes, such as predictive differentiation, may depend on other hippocampal-region structures, such as the dentate gyrus and hippocampus. These functions would therefore be eliminated by a selective H-lesion in the model.

We will return later to the issue of which hippocampal-region structures may be involved in these other functions. For now, we expect the new



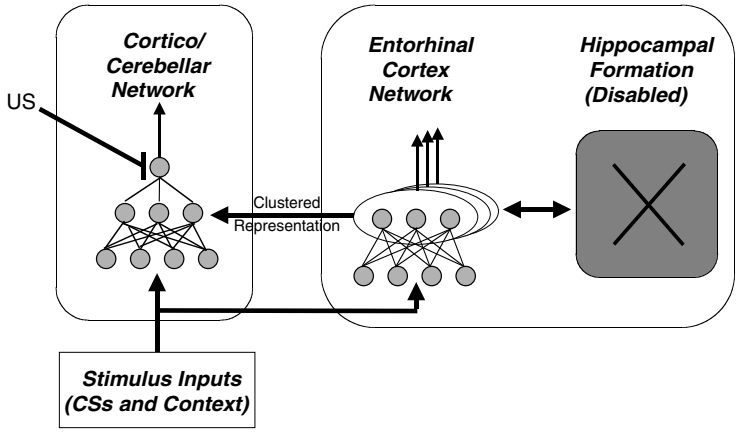


Figure 9.7 (A) The intact cortico-hippocampal model assumes that the operation of all hippocampal-region structures can be approximated by a single hippocampal-region network that produces new stimulus representations that are biased by redundancy compression and predictive differentiation. These new representations are then provided to the cortico/cerebellar network, which learns to map from them to a behavioral response. (B) Broad HR lesion is simulated by disabling all hippocampal-region processing; only the cortico/cerebellar network is assumed to operate. (C) A selective lesion of the hippocampus (the H lesion) may spare entorhinal processing, since the entorhinal cortex has reciprocal connections with other cortical areas that provide its major sensory input and receive its output (refer to figure 9.1). The H-lesion model assumes that only the entorhinal network is available to modify stimulus representations and provide these to the cortico/cerebellar network.

H-lesion model to be sufficient to show just those hippocampal-region-dependent behaviors that depend on redundancy compression but not those that depend on predictive differentiation. This model predicts that there should be a similar pattern of effects following selective H-lesion in animals.

Behavioral Predictions of the H-Lesion Model. In chapter 7 we discussed how hippocampal-region damage disrupts **latent inhibition,** a behavioral paradigm in which learning a CS-US association is slower in animals that were given prior exposure to the CS (CS Exposure condition) relative to animals that were given equivalent exposure to the context alone (Sit Exposure condition). Hippocampal-region damage disrupts latent inhibition, so animals in the CS Exposure condition learn as fast as animals in the Sit Exposure condition.

In chapter 7, we simplified the story slightly: Although broad damage to the hippocampal region does indeed abolish latent inhibition, a more precise lesion that is limited to the hippocampus but spares the entorhinal cortex may spare or even enhance latent inhibition.*[8]

Furthermore, lesions that are limited to the entorhinal cortex (and possibly subiculum) but that spare hippocampus and dentate gyrus do suffice to disrupt latent inhibition (figure 9.8A).[9] Thus, it seems that while the hippocampus (and possibly dentate gyrus) are not necessary for latent inhibition, the entorhinal cortex is.

This finding is exactly as predicted by our cortico–hippocampal model. Recall from chapter 7 that the cortico-hippocampal model explains latent inhibition in terms of redundancy compression: During exposure, the representation of the CS is compressed together with the representation of the context in which it occurs, since neither predicts any US. Later, when the task is to respond to the CS but not the context alone, this compression must be explicitly undone, and learning is slowed while this redifferentiation takes

*One important exception to this rule is a study in which an ibotenic acid lesion that was limited to hippocampus and dentate gyrus impaired latent inhibition in rat appetitive conditioning; that is, lesioned rats learned equally quickly, whether or not they had previously been exposed to the CS (Han, Gallagher, & Holland, 1995). The reasons for this anomalous result are still unclear; one factor may be that Han, Gallagher, and Holland used a procedure in which animals were first exposed to the CS, then trained to approach the food cup for a reward, then trained that the CS predicted this reward. This paradigm contrasts with the more standard sequence in which animals are first trained to obtain food, then exposed, then given CS-US pairings (e.g., Honey & Good, 1993). Having a training session interposed between CS exposure and CS-US pairing may disrupt the effect, particularly in animals with hippocampal lesion, which may quickly forget the earlier exposure.

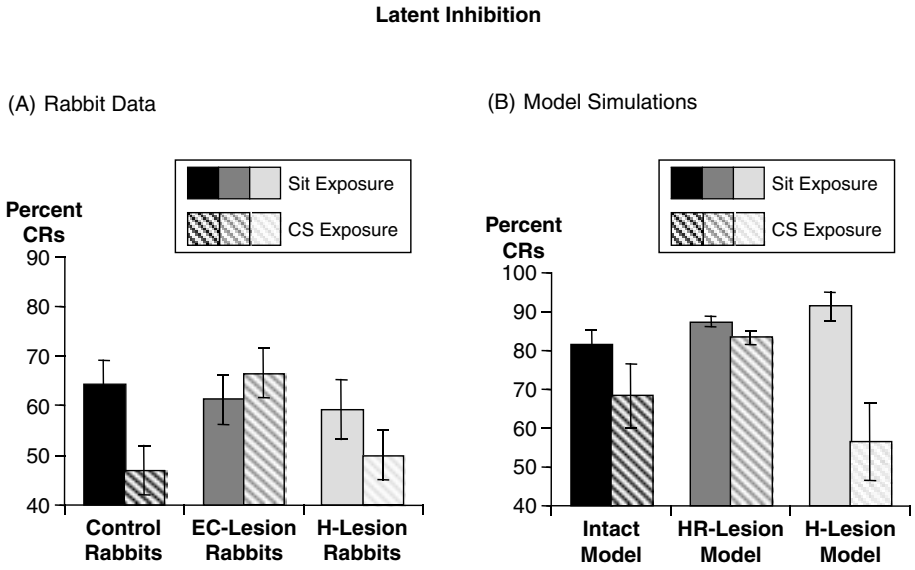


Figure 9.8 Latent inhibition: Learning a CS-US association is retarded in a group given prior CS exposure compared with a group given equivalent exposure to the context alone (Sit Exposure). (A) Latent inhibition is eliminated by lesion that includes entorhinal cortex (ECL) but not by a selective hippocampal lesion (HL) that spares entorhinal cortex. (Data from Shohamy, Allen, & Gluck, 1999.) (B) The cortico-hippocampal model shows the same behavior: The H-lesion model but not the HR-lesion model shows latent inhibition. Note that in these and subsequent figures, the intact and HR-lesion models were simulated under slightly different conditions than the H-lesion model; for this reason, it is not appropriate to compare learning times, etc., between models (e.g., HR-lesion versus H-lesion), but only to compare a model's performance under one training condition with its own performance on a different training condition (e.g., exposed versus nonexposed).

place. Since this explanation of latent inhibition depends wholly on redundancy compression, it can be mediated by the entorhinal network alone; thus, the H-lesion model does show latent inhibition (figure 9.8B), though the HR-lesion model does not.[10]

A related effect is **learned irrelevance,** in which prior exposure to a CS *and* a US, uncorrelated with each other, slows subsequent learning of a CS-US association.[11] Again, our cortico-hippocampal model provides an account of this phenomenon in normal animals: During exposure, neither the context nor the CS is a good predictor of the US, so they become compressed together. As in latent inhibition, this compression hinders subsequent learning to associate the CS, but not the context alone, with the US.[12] Again, the model predicts that the learned irrelevance effect depends primarily on

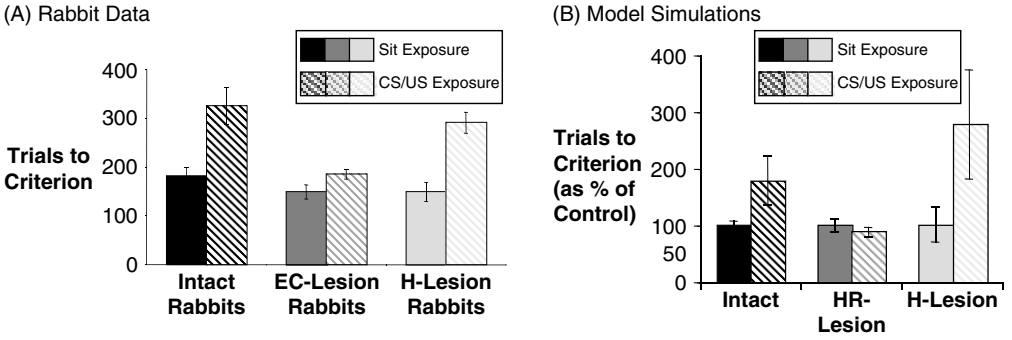


Figure 9.9 Learned irrelevance (retarded CS-US learning after exposure to CS and US, uncorrelated with each other). (A) Learned irrelevance is eliminated by lesion that includes entorhinal cortex (EC lesion) but not by a selective hippocampal lesion (H lesion) that spares entorhinal cortex. (Data from Allen, Chelius, & Gluck, 1998.) (B) The cortico-hippocampal model shows the same behavior: The H-lesion model but not the HR-lesion model shows learned irrelevance.

redundancy compression, and therefore the entorhinal cortex should be able to mediate learned irrelevance. In other words, HR lesion should disrupt learned irrelevance but H lesion might not, as is shown in figure 9.9B. In recent studies of rabbit eyeblink conditioning in our laboratory, we have now confirmed this prediction, as is shown in figure 9.9A.[13]

Several other novel predictions arise from our cortico-hippocampal model. For example, in **sensory preconditioning,** prior exposure to a compound AB increases the degree to which subsequent learning about A generalizes to B.[14] The intact model, but not the HR-lesioned cortico-hippocampal model, shows sensory preconditioning (figure 9.10A). The cortico-hippocampal model explains this effect in terms of redundancy compression: The representations of A and B are compressed during exposure to the compound, which increases later generalization between the components. Since this explanation involves redundancy compression, our model predicts that sensory preconditioning should survive selective H-lesion that spares the entorhinal cortex.[15] This is a prediction that remains to be tested in animals.

Easy-hard transfer of learning occurs when prior training on an easy discrimination facilitates learning a harder discrimination along the same stimulus continuum. For example, phase 1 might involve learning to discriminate a very low tone (S1+) from a very high tone (S4−). Phase 2 would involve transfer to a new discrimination along the same stimulus continuum—tone frequency in this case—such as discriminating a medium low tone (S2+) from a medium high tone (S3−). Prior training on the "easy" discrimination in

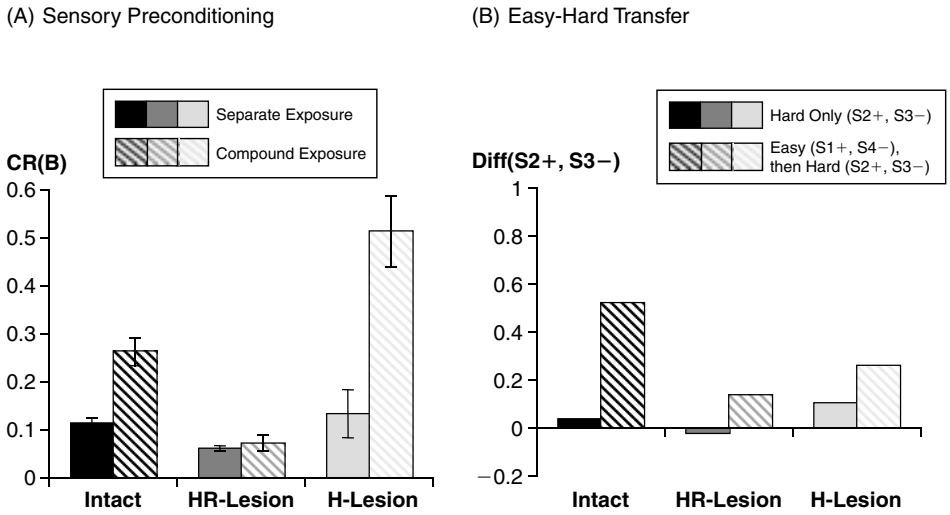


Figure 9.10 (A) Sensory preconditioning in the cortico-hippocampal model. In the intact model, exposure to the compound AB followed by A-US training leads to increased responding to B, compared with a control condition involving separate exposure to A and B. This sensory preconditioning effect is maintained in the H-lesion model but not in the HR-lesion model. (B) Easy-hard transfer (learning a hard discrimination, S2+, S3−, is facilitated by prior training on an easy discrimination, S1+, S4−, along the same stimulus continuum—more than equivalent amounts of prior training on the hard discrimination alone). The intact cortico-hippocampal model shows a strong transfer effect; after a fixed number of trials on the hard discrimination, the difference in responding to S2+ and S3− (Diff(S2+, S3−)) is stronger after prior training on the easy discrimination than after prior training on the hard discrimination itself. The HR-lesion model shows only a small (nonsignificant) easy-hard effect, which may be attributed simply to stimulus generalization. (The difference between the intact model and the HR-lesion model performance reflects the effect of representational changes in the intact model.) Similarly, the H-lesion model shows only a weak (nonsignificant) effect of prior easy training, again attributable to stimulus generalization.

phase 1 can often speed subsequent learning of the "hard" discrimination in phase 2—more than an equivalent amount of pretraining on the hard discrimination itself.[16]

We have hypothesized that this kind of easy-hard transfer would depend on predictive differentiation.[17] Learning about the easy discrimination in phase 1 results in differentiation of the representations of those stimuli, as more and more resources are devoted to encoding the relevant features. This differentia-

tion will make all stimuli along the same continuum easier to distinguish, facilitating the hard task. As a result, the intact cortico-hippocampal model shows faster learning of the hard discrimination following prior training on the easy discrimination compared with a control condition given training on the hard discrimination all along, as is shown in figure 9.10B.[18]

The HR-lesion model shows only a minimal effect, and this is due to stimulus generalization whereby a response learned to S1+ generalizes partially to the nearby S2+. In contrast, the difference between the performance of intact and HR-lesion models in the easy-hard condition reflects the effect of representational changes in the intact model. Because the easy-hard effect is explained in terms of representational differentiation, rather than compression, the entorhinal network is not sufficient to mediate the effect. Thus, the H-lesion model does not show easy-hard transfer.[19] This prediction of the model remains to be tested in animals.

In sum, our model predicts that behaviors that depend on hippocampal-region mediation can be distinguished according to whether they primarily reflect representational compression, predictive differentiation, or both. Since the model expects that the entorhinal cortex is sufficient to mediate representational compression, behaviors that primarily reflect this process should survive selective hippocampal lesion. In contrast, behaviors that involve predictive differentiation may depend on other structures, such as the hippocampus and dentate gyrus, and may be eliminated following a selective hippocampal-region lesion.

Contextual Processing After H Lesion. Chapter 7 focused on another class of behaviors that depend on the hippocampal region: contextual processing. Recall that after normal animals are trained to respond to a conditioned stimulus cue (e.g., tone) in a particular context X, there may be decreased responding when the tone is presented in a new context Y.[20] The intact cortico-hippocampal model shows a similar response decrement after context shift, as is seen in figure 9.11. In the intact model, the hippocampal-region network creates a new representation of the tone that includes information about the context X in which the tone occurs. This representation is acquired by the cortico/cerebellar network and then mapped to a behavioral response. When the tone is presented in a new context Y, the representation of tone-in-X is only partially activated, and so the response is only partially activated. This explanation assumes that the context-shift effect depends on hippocampal-region representational processes, and so our HR-lesion model does not show the effect.

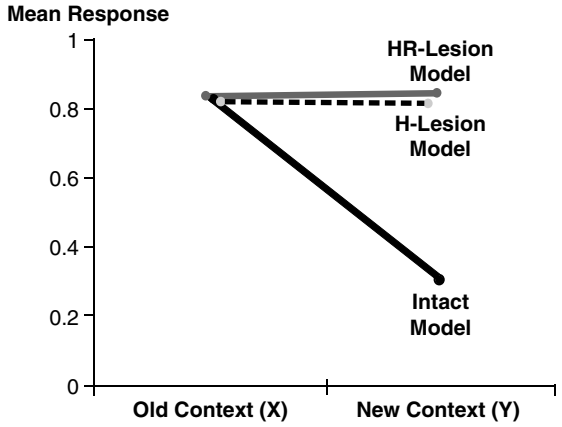


Figure 9.11 After training to respond to a cue A in a context X, the intact model but not the HR-lesion model shows a decrement in responding when A is presented in a new context Y. This is consistent with animal behavior (refer to figure 7.11A). The H-lesion model does not show a response decrement after context shift. Similarly, selective H lesion abolishes the context shift effect in rats (Honey & Good, 1993). (Adapted from Myers & Gluck, 1994, figure 3A, and Myers, Gluck, & Granger, 1995, figure 9.)

What about the H-lesioned model? The H-lesioned model does include representational compression in the entorhinal network, and as the tone is repeatedly presented in context X, the representations of tone and X should be compressed. Thus, at first glance, it might seem that the context shift effect should be mediated by entorhinal compression and should survive selective hippocampal lesion. However, the H-lesioned model does not include representational differentiation, a process that allows the representation of tone-in-X to be differentiated from that of X alone in the intact model. This makes all the difference.

For example, consider the simplified picture of the entorhinal network shown in figure 9.12. This network has four inputs: one representing the presence or absence of tone and three representing various features of the context. There are three patches of three nodes, each of which responds to this input. When context X is presented alone (figure 9.12A), one node in each patch wins the competition and becomes active. When the tone is presented in X (figure 9.12B), the input has considerable overlap with that encoding the context X alone; many of the clusters respond the same way to both. In this example, only one patch (the leftmost one) responds differently to X and tone-in-X.

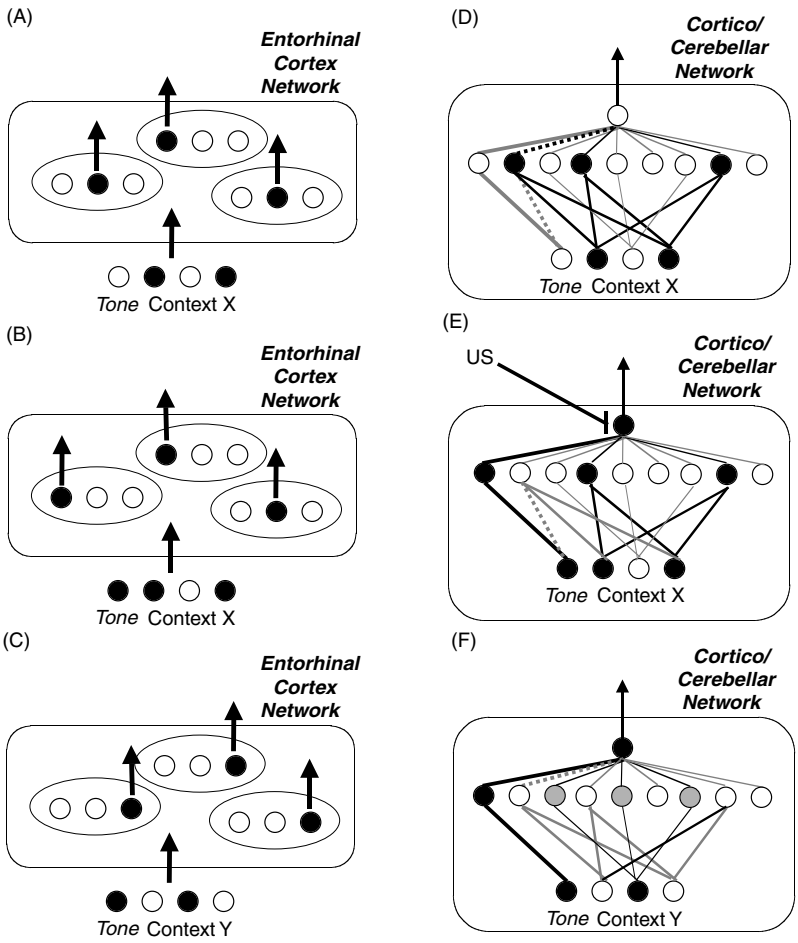


Figure 9.12 Loss of contextual sensitivity in the H-lesion model. For example, assume that the entorhinal network consists of nine nodes organized into three clusters. During tone+ training in context X, tone-in-X+ and X− trials are intermixed. Because of representational compression in the entorhinal network, the responses to tone-in-X+ and X− are liable to become similar. Thus, X− evokes one set of responses from the cluster (A), while tone-in-X+ evokes a very similar response (B): Only one cluster outputs differently to tone-in-X+ (the leftmost cluster). The cortico/cerebellar model adopts these entorhinal representations and learns to map from them to the correct response. This can be done only by learning strong weights from the elements that are present in the entorhinal response to tone-in-X+ but not to X−: Namely, the leftmost internal-layer node (activated by the tone) has a strong positive connection to the output (E), while the neighboring internal-layer node (active when the tone is absent) has a strong inhibitory connection to the output (D). Now the response of the network depends wholly on whether the tone is present; the internal-layer nodes encoding context information are ignored at the output level. When the tone is presented in a very different context Y, the entorhinal network will generate a very different pattern of responses (C), but these are adapted by the cortico/cerebellar network only over a series of many trials. On the initial presentation, the presence of the tone in the input is enough to drive a strong response (F). Hence, the H-lesion model continues to generate a strong response when the tone is presented in a new context: The context-shift effect is eliminated.

In the H-lesioned model, there is no hippocampal-mediated differentiation mechanism to pull these representations of X and tone-in-X apart. They are passed as is to the cortico/cerebellar model. In this simple example, there are nine internal-layer nodes in the cortico/cerebellar model—the same as the number of nodes in the entorhinal network—and so the internal-layer nodes eventually come to exactly mimic the entorhinal activations. Thus, the internal-layer activations for X alone (figure 9.12C) are the same as the entorhinal activation for X alone (figure 9.12A), and the internal-layer activations for tone-in-X (figure 9.12D) are the same as the entorhinal outputs for tone-in-X (figure 9.12B)

Now the cortico/cerebellar network must learn to map from these representations to the correct responses. Since there is such a high overlap between the representations of tone-in-X and context X alone, the cortico/cerebellar network must assign high weights to exactly those internal-layer nodes that are different between the two representations. In this case, that is only the leftmost internal-layer node (which responds to tone-in-X but not X alone) and its neighbor (which responds to X alone but not to tone-in-X). By assigning a strong positive weight to the leftmost node (and a strong negative weight to its neighbor), the network can learn to generate a behavioral response to tone-in-X but not to X alone. In effect, the cortico/cerebellar network of the H-lesion model learns to ignore the context and produce a response that depends only on whether the tone is present or absent.

Now suppose that the same tone is presented in a novel context Y that is very different from X. The entorhinal network will generate a new set of responses, as is illustrated in figure 9.12E, but these will be adopted by the cortico/cerebellar network only over the course of many repeated trials. On the very first presentation of tone-in-Y, Y will activate some subset of internal-layer nodes in the cortico/cerebellar network, and by virtue of prior training, the tone will continue to activate the leftmost internal-layer node, as is seen in figure 9.12F. But this leftmost node has a strong positive connection to the output, and so there will be a strong behavioral response—even in the new context Y, as is shown in figure 9.11.

The H-lesion model therefore predicts that selective hippocampal lesion that spares the entorhinal cortex should disrupt this context-shift effect.[21] Indeed, rats with selective hippocampal lesions do show no decrement in responding with context shift,[22] just as is expected by our model.

This finding of no context-shift effect in the H-lesioned model is particularly interesting because it is a model prediction that was not immediately obvious from the qualitative statement of theory. Only by implementing and observing the model in action does the full range of predictions become

apparent. This is one reason why computational modeling is often valuable: It can help to make explicit the subtle implications of a theory.

The H-lesion model is similarly insensitive to contextual effects in latent inhibition. Recall that, in intact animals (and the intact cortico-hippocampal model), latent inhibition is disrupted if there is a context shift between exposure and training phases (figure 7.11). The H-lesion model shows latent inhibition (figure 9.8B) but does not show disrupted latent inhibition after context shift. Once again, exposure results in a compression of a stimulus cue with the context in which it occurs; this retards subsequent learning to respond to the cue but not to the context alone. Exposure also results in an overall reduction of the representational space allocated to the cue. Shifting to a new context for training can alleviate the former but not the latter. Thus, the H-lesion model continues to show latent inhibition even in a new context.[23] This is consistent with the finding that selective hippocampal lesions reduce the context sensitivity of latent inhibition.[24]

In general, the H-lesion model predicts that selective H lesion will greatly disrupt the contextual sensitivity of learned associations—just as HR lesion does. This prediction is consistent with two existing studies, as was mentioned above, but further empirical studies are needed to test this prediction fully.

Other Hippocampal-Region Subfunctions. At this point, the entorhinal network and the H-lesion model that incorporates it are still largely speculative. The entorhinal network appears to be compatible with known features of the anatomy and physiology of entorhinal cortex, and the H-lesion model appears to account for what data exist regarding the effects of selective hippocampal lesion on classical conditioning. However, these data are relatively scant. There is clearly much more empirical work to be done—for example, testing the predictions described early for lesion studies of sensory preconditioning and easy-hard transfer. Our model also predicts that there should be a distinction between multimodal compression, which ought to depend on entorhinal cortex, and unimodal compression, which might be done in sensory cortices and thus survive entorhinal lesion. To date, there is little information comparing multimodal and unimodal stimulus processing in H-lesioned animals.

Our H-lesion model also begs an important question: If the entorhinal cortex is performing redundancy compression among co-occurring stimuli, where are the remaining postulated hippocampal-region functions localized? The original specification of our cortico-hippocampal model proposed that the hippocampal region was involved in compressing representations of stimuli that co-occur or have similar meaning while

Table 9.1 Summary of Representational Biases Assumed by the Cortico-Hippocampal Model to Reflect Hippocampal-Region Mediation

	Bias to Compress	Bias to Differentiate
Stimulus-Stimulus Relationships	Compress representations of co-occurring stimuli	Differentiate representations of stimuli that don't co-occur
Stimulus-Outcome Relationships	Compress representations of stimuli that predict similar outcomes	Differentiate representations of stimuli that predict different outcomes

differentiating the representations of stimuli that do not co-occur or that have different meanings.[25] These four representational constraints are summarized in table 9.1.

The entorhinal cortex can perform representational compression of co-occurring stimuli, termed stimulus-stimulus compression in table 9.1. Another kind of representational compression involves stimuli that have the same meaning. For example, if two stimuli, such as a tone and a light, both reliably predict a food US, then our theory argues that the representations of the tone and the light should be compressed, so that subsequent learning about the tone will generalize to the light. In fact, such an effect is seen in normal animals: If a tone is subsequently paired with a shock US, animals will tend to generalize this new learning to the light as well. This effect is termed **acquired equivalence.**[26] It is still an open question whether HR-lesion impairs acquired equivalence, but the cortico-hippocampal model predicts that it should.

Differentiation and the Dentate Gyrus. The other type of representational recoding that occurs in our intact cortico-hippocampal model is representational differentiation. Differentiation requires decreasing the overlap in representation between two stimuli so that they never co-occur or so that they predict different future reinforcement. The easiest way to differentiate representations is to ensure that any nodes (or neurons) that respond to one stimulus do not respond to the other and vice versa. Several researchers have proposed that the dentate gyrus, which lies between entorhinal cortex and hippocampus in the processing chain of figure 9.1, could perform such a function.[27] The first reason to suspect that the dentate gyrus may perform differentiation of representations is anatomical. In the rat, there are about 100,000 pyramidal cells in entorhinal cortex, and these project to about one million neurons in dentate gyrus (called **granule cells**). This means that

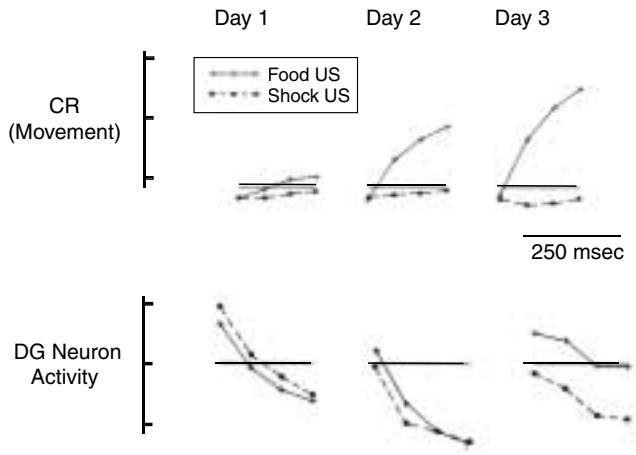


Figure 9.13 Neurons in the dentate gyrus come to differentiate stimuli that are mapped to different responses. Each graph shows the average response pattern to each CS on a given day of training. Top row: Rats are trained that one tone CS predicts a food US while another tone CS predicts a shock US. On day 1 of training, rats respond similarly to both; by days 2 and 3, rats show head movement responses to the CS that predicts food but not to the CS that predicts shock. Bottom row: activity recorded from cells in the dentate gyrus. On day 1, there is a brief increase in responding when either CS is presented, which lasts for about 250–500 msec and is followed by below-baseline activity. By day 3 (after the behavioral response has been acquired), there is a differential response to the CSs: excitation in response to the CS associated with food and inhibition in response to the CS associated with shock. (Adapted from Segal & Olds, 1973, figure 3.)

information fans out in a ratio of about 1:10 as it travels from entorhinal cortex to dentate gyrus, and this alone could help to increase the difference between stimulus items because more neurons yield more possible patterns.*[28]

Thus, the anatomy of the dentate gyrus would be consistent with automatically differentiating the representations of any random inputs. But there is also evidence that the dentate gyrus actively differentiates inputs that predict different outcomes.

In one study by Segal and Olds, rats were trained that one tone predicted a food US while a different tone predicted a shock US.[29] Figure 9.13 (top row)

*Another clue that the dentate gyrus might participate in representational differentiation comes from physiology: In the dentate gyrus, granule cells fire very infrequently; for any single stimulus, only a small percentage of granule cells respond (Treves & Rolls, 1992, 1994; Jones, 1993). This would be consistent with the idea that the dentate gyrus is orthogonalizing inputs, meaning that it reduces the representational overlap.

shows that on day 1 of training, rats responded to neither tone. However, by days 2 and 3, rats were showing reliable head-movement responses to the tone that predicted food but not to the tone that predicted shock. During learning, the experimenters recorded granule cell activity in the dentate gyrus. Figure 9.13 (bottom row) shows that on day 1 of training, the granule cells exhibit a short (250–500 msec) excitatory response to either CS, followed by a short (250–500 msec) inhibitory response. This pattern of responding was sensory-evoked, meaning that it depended on presentation of a tone CS rather than on that tone's meaning. By day 3, however, the response changed, and granule cells gave differential responses to the two tones: The positive CS (associated with food) evoked a short excitatory response, while the negative CS (associated with shock) evoked a short inhibitory response. This differentiated response occurred only after the behavioral response was well learned, and it did not occur unless the CSs were associated with different outcomes.[30]

In a related experiment by Deadwyler and colleagues, rats were again trained that one cue (e.g., tone) predicted a US while a second (e.g., light) did not.[31] Again, the dentate granule cells developed differential responses to the two cues, responding to the tone but not the light. Next, the experimenters reversed the contingencies so that the light but not the tone predicted the US. Gradually, the granule cells reversed their activity to reflect the new situation, developing responses to the light but not the tone.

Perhaps most significant of all, these differential responses to the two CSs were *not* visible in the entorhinal cortex, from which dentate gyrus receives most of its input regarding sensory stimuli.[32] This implies that the patterns of activity in the dentate gyrus were not just copies of distinctions made elsewhere in the brain, but that the dentate gyrus itself is integrally involved in differentiating the representations of stimuli that make different predictions about upcoming reinforcement.

Thus, in the same way that entorhinal cortex may perform representational compression, there is emerging evidence that dentate gyrus may perform representational differentiation. These two structures together could perform most or all of the representational processing that we have assumed takes place in the hippocampal region. Other hippocampal-region areas, such as CA3 and CA1, may be more integrally involved in short-term storage of this information and overseeing its eventual consolidation to cerebral cortex. A major focus of current and future modeling work will be expanding the cortico-hippocampal model to include both entorhinal and dentate (and CA3 and CA1) components and studying the manner in which they could interact to provide a more complete account of hippocampal-region function.

Stimulus Competition in Entorhinal Cortex

A very different conception of entorhinal cortex has been proposed by Nestor Schmajuk and colleagues. Figure 9.14A contains the simplified drawing of the Schmajuk-DiCarlo (S-D) model discussed in chapter 6.[33] As a brief review, the basic idea in this model is that the hippocampal region has two

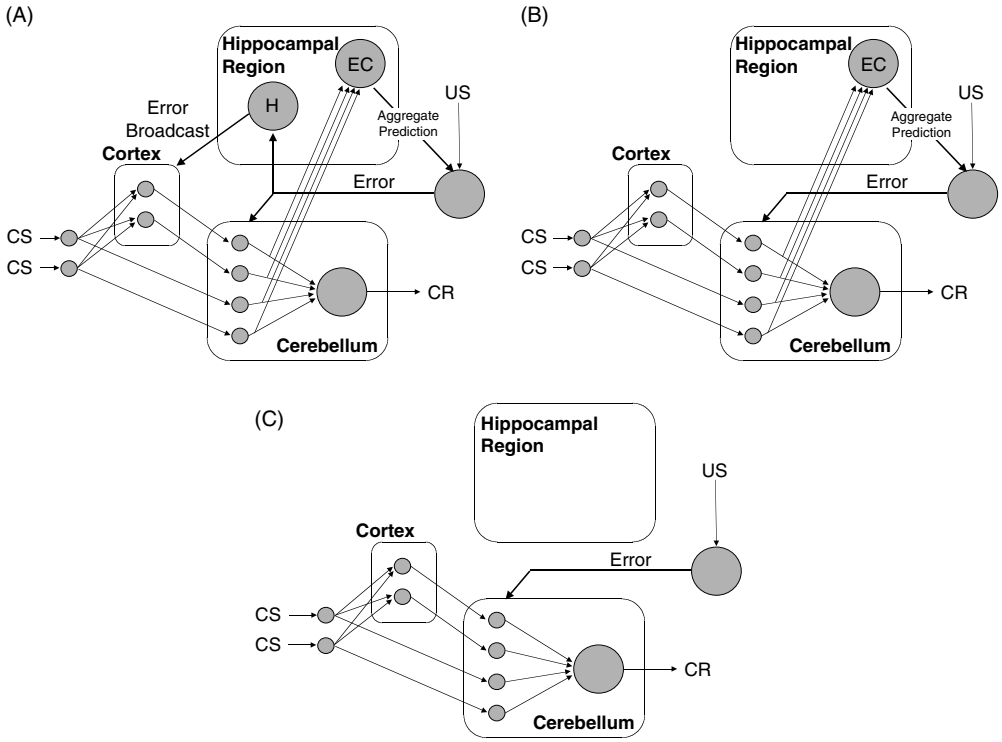


Figure 9.14 (A) The Schmajuk-DiCarlo (1992) model assumes that the hippocampal region has two basic functions. First, it is responsible for calculating the aggregate prediction of the US. This information is used to calculate overall prediction error (the difference between aggregate prediction and actual US). Second, the hippocampal region is assumed to be responsible for broadcasting this prediction error to the cortex. It is in the cortex that nodes learn to respond to configurations of stimuli. Later, Buhusi and Schmajuk (1996) proposed that the aggregate prediction could be mapped onto the entorhinal cortex and the error broadcast onto the hippocampus. (B) Selective H lesion would therefore disrupt learning in the cortex but not in the cerebellum. The resulting network could still learn direct CS-US associations but not configural CS-CS associations. (C) Broad HR lesion would disrupt both functions, disrupting configural learning in the cortex and also stimulus competition in the cerebellum. Direct CS-US associations could still be learned, but stimulus competition effects such as blocking would be disrupted.

functions. First, it is responsible for calculating the aggregate prediction of the US: On the basis of all available information (i.e., what CSs are present), how strongly is a US expected? The difference between this aggregate prediction and the actual US is the prediction error. Their second putative function for the hippocampal region is to broadcast this prediction error back to the cortex.

Hippocampal-region damage is assumed to damage both the aggregate prediction and error-broadcast functions (figure 9.14C). Without the aggregate prediction, the system no longer performs functions such as blocking that depend on how well the US is predicted by all available cues; without the error broadcast, the configural nodes in the cortex cannot be trained, and so learning is limited to direct CS-cerebellar projections.

In later work, Schmajuk and colleagues have considered how various components of their models can be mapped onto individual brain structures.[34] Specifically, Schmajuk and colleagues have suggested that the hippocampus might be principally involved in the error broadcast function, while the entorhinal cortex might compute the aggregate prediction. Thus, a selective lesion of hippocampus that spared entorhinal cortex might eliminate the ability to learn about cue configurations (figure 9.14B), while a broader HR lesion that included entorhinal cortex might also eliminate the ability to do cue competition (figure 9.14C).

Schmajuk and colleagues' proposal that the entorhinal cortex mediates cue competition is very different from the implications of our own proposal that the entorhinal cortex performs redundancy compression. As a result, the two models make some opposing predictions about the effects of selective H lesion. One example is sensory preconditioning. Figure 9.10A illustrated the prediction of the cortico-hippocampal model that selective H lesion should spare sensory preconditioning. The Schmajuk et al. model makes the opposite prediction. In this model, phase 1 exposure to a stimulus compound AB normally results in the formation of a configural node in the cortex that responds to this configuration (figure 9.15A). Next, in phase 2, A is paired with the US. Most associative weight accrues to a direct connection between A and the US (figure 9.15B), but A also partially activates the AB configural node, and so the weight from AB to US is also partially strengthened. Finally, when B is presented in phase 3 (figure 9.15C), it partially activates the AB node, which in turn activates a prediction of the US. Thus, in the Schmajuk model, sensory preconditioning depends on configural learning. According to the model, configural learning in turn depends on error broadcasts by the hippocampus. If this is removed, then there is no way to form configural nodes in the cortex (figure 9.15D). The model can still learn that A predicts

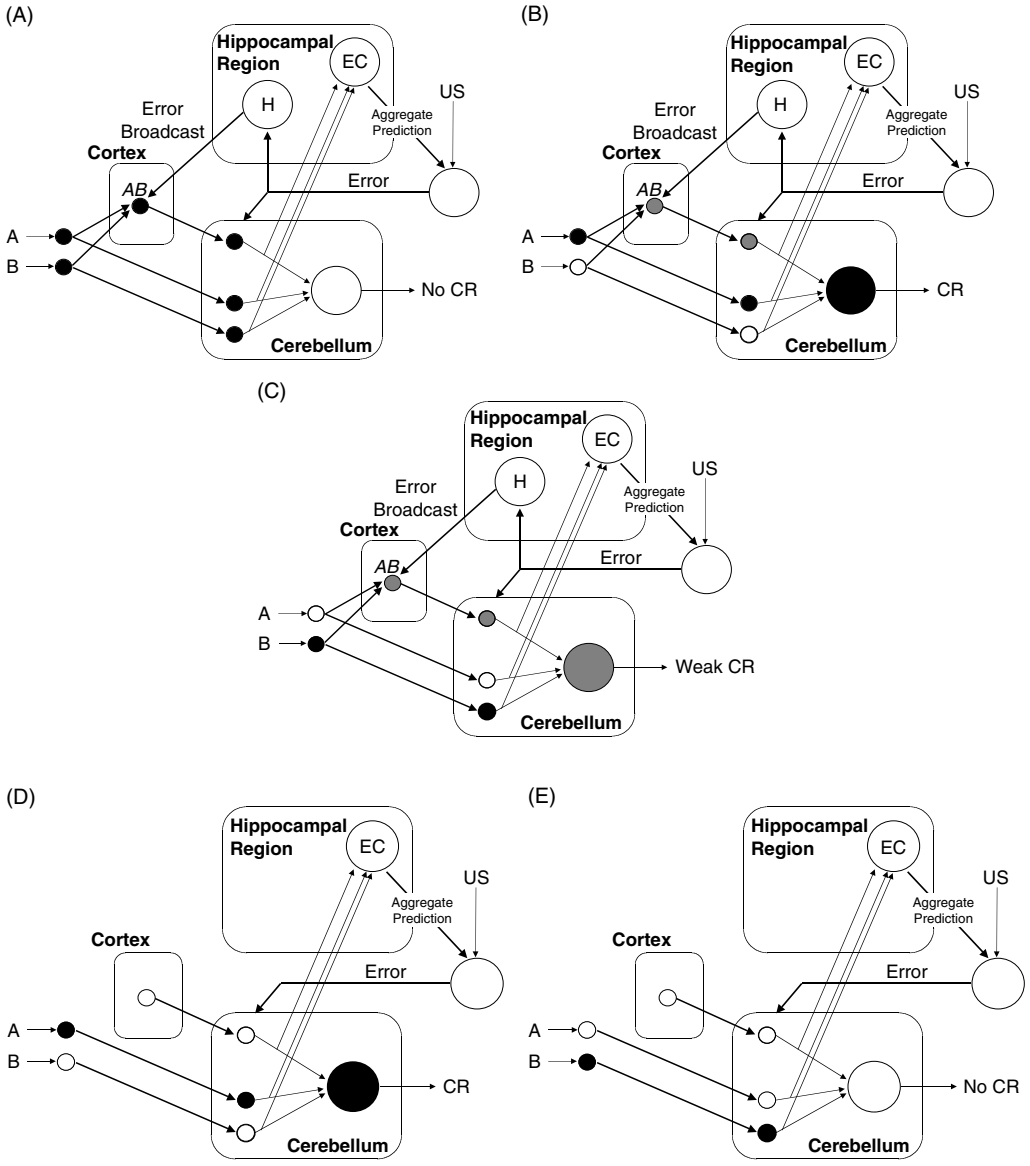


Figure 9.15 Sensory preconditioning in the intact model of Schmajuk and DiCarlo (1992). (A) In phase 1, CSs A and B are presented together. A configural node forms in cortex to encode the AB pairing; this and the components both project to the cerebellum, where they are not associated with any response. (B) In phase 2, A is paired with the US. The AB configural node is partially activated and acquires some association with the US. (C) In phase 3, B is presented alone, partially activates the configural AB node, and evokes a weak CR. (D) Selective H lesion eliminates the hippocampal error signal, meaning that the cortex cannot form new configural nodes. Phase 1 exposure to the compound AB does not result in formation of an AB node in cortex. In phase 2, only the direct A-US association is learned. (E) When B is presented alone, little or no response is evoked.

the US in phase 2, but when B is presented (figure 9.15E), little response is evoked.

Thus, the S-D model predicts that selective H lesion should disrupt sensory preconditioning.[35] This is the opposite prediction from our cortico-hippocampal model. Empirical studies are needed to determine which model's predictions are accurate on this issue. Currently, the only available data show that both fimbrial lesions and kainic acid injections into hippocampal field CA1 abolish sensory preconditioning. However, neither of these lesions provides the sort of precise hippocampal lesions that are created by ibotenic acid, as illustrated in figure 9.4B.[36]

Another interesting feature of the Schmajuk et al. model is that it contains several separate substrates that all contribute to a latent inhibition effect. Thus, different kinds of hippocampal-region lesions may have inhibitory effects, no effect, or even facilitatory effects on latent inhibition.[37] These arguments depend on extrahippocampal structures that are beyond the scope of the simple illustration in figure 9.15. Nonetheless, the model provides a large set of predictions that should keep empirical researchers busy for some time.

Backprojections from Entorhinal Cortex

The previous two sections described computational models that consider how the entorhinal cortex could operate on stimulus inputs. Both our cortico-hippocampal model and the models of Schmajuk and colleagues assume that the hippocampal-region operates on sensory information and then transmits its output back to cortex. In our cortico-hippocampal model, this takes the form of stimulus representations to be acquired by cortex. In the models of Schmajuk and colleagues, this takes the form of an error broadcast from the hippocampal region to cortex. In either case, an important question is to consider projections from the hippocampal region—specifically, from entorhinal cortex—back to other cortical areas.

Figure 9.16 shows a schematized drawing of cortico-cortical information flow, including both feedforward (solid) and feedback (dotted) pathways.* Sensory information reaches pyramidal neurons in the neocortex: first primary sensory cortex, then higher-order sensory cortex, and eventually polymodal association areas. This information contacts dendrites of pyramidal neurons in both the superficial and deep layers of cortex; neurons in the

*This simplified drawing, which follows Rolls, 1996, illustrates only the major projections under discussion. Additional neuron types and intrinsic projections are not shown.

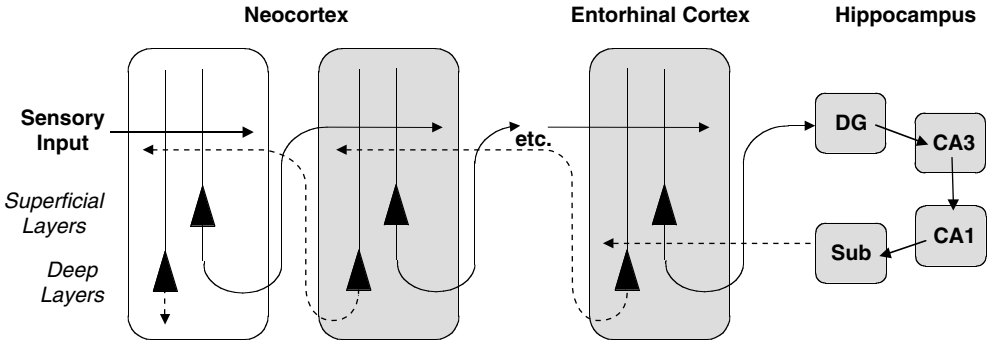


Figure 9.16 Schematic of sensory information flow through cortical areas as envisioned by Rolls (1996). Sensory inputs from previous cortical processing stages contact dendrites of pyramidal cells in superficial and deep layers of neocortical association areas. Axons from neurons in the superficial layers provide output that travels to subsequent cortical processing areas. This *feedforward* process is repeated through various neocortical sensory and association areas until information reaches the entorhinal cortex and then the dentate gyrus, hippocampus, and subiculum (additional intrinsic connections not shown). Information returns through a *feedback* pathway to deep layers of entorhinal cortex (dotted line) and then back to areas of cortex representing progressively earlier processing stages. DG = dentate gyrus; Sub = subiculum.

pyramidal layers continue the feedforward pathway, sending output on to other cortical areas that represent subsequent processing stages. Eventually, the information reaches entorhinal cortex and, from there, dentate gyrus, hippocampus, and subiculum. From there, information is projected back to cells in the cortical areas associated with earlier processing stages. This appears to be an anatomical pathway by which hippocampal-region processing can influence cortical areas.

Edmund Rolls has developed a theory of how these feedback projections could allow representations developed in hippocampus to drive cortical storage.[38]

Events might occur along the lines shown in figure 9.17. Pyramidal neurons in an area of cortex receive feedforward projections carrying sensory information from other (earlier) cortical areas and also feedback projections from other (later) cortical areas (figure 9.17A). In turn, they themselves send output on to later cortical areas and back to earlier cortical areas. In the case of entorhinal cortex, for example, feedforward inputs detail the highly processed, multimodal features of current input (vision and taste are shown in figure 9.17A), while feedback inputs carry information about the hippocampal region's rerepresentation of this information. Initially, these inputs

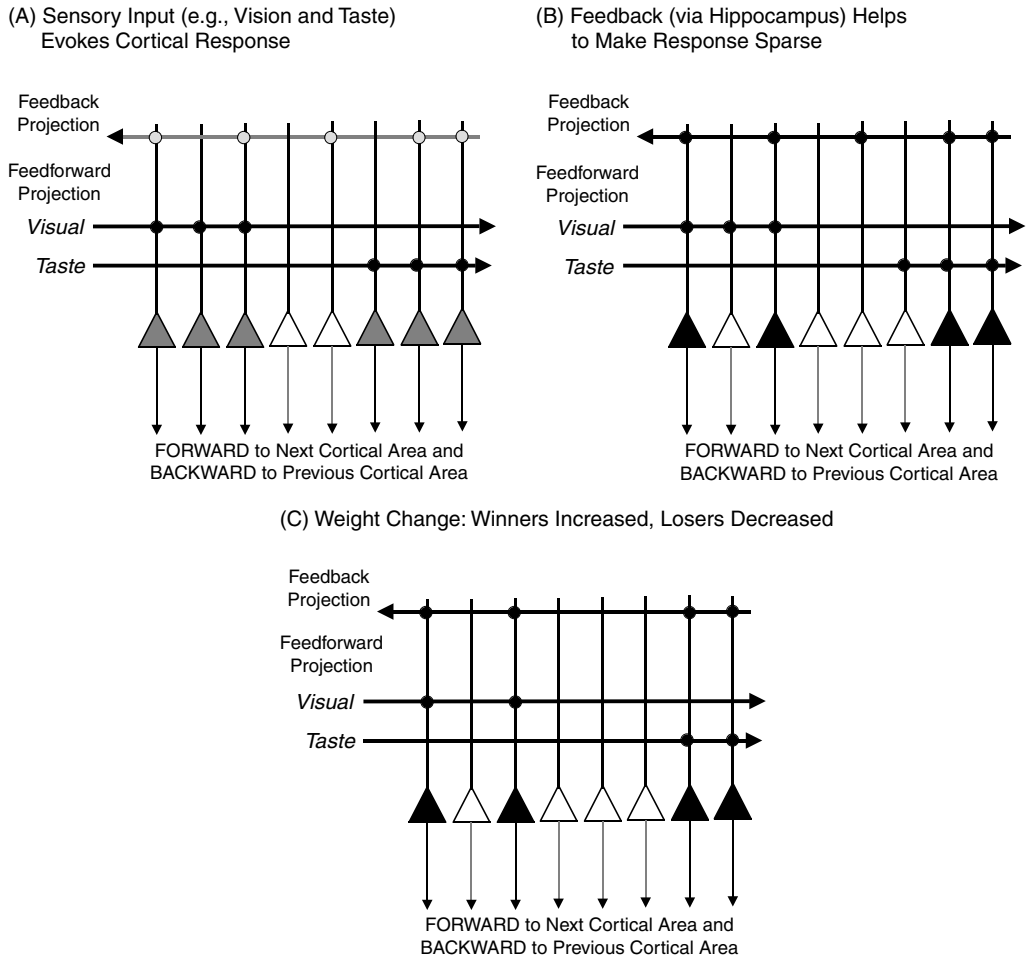


Figure 9.17 How cortical backprojections might transfer hippocampal representations to cortex, along the lines proposed by Rolls (1989, 1996). (A) Pyramidal neurons in an area of cortex receive both feedforward projections from earlier cortical processing areas and feedback projections from later cortical processing areas. Neurons that receive feedforward inputs become partially active and project to subsequent processing areas. In the case of the entorhinal cortex, for example, feedforward inputs specify highly processed multimodal features of stimuli, such as sight and taste; feedback inputs specify hippocampal-region representations of these inputs. (B) The hippocampal region provides the last processing step in the chain, and information begins to flow backward. Where feedforward and feedback projections converge, neurons are strongly activated. In general, the hippocampal-mediated feedback tends to make cortical activity sparse,[3] reducing the number of active neurons. (C) Following the rules of a competitive network, cortex undergoes plasticity: Neurons that are most active win, and their weights from active inputs are strengthened. Weights to nonwinning nodes are weakened.

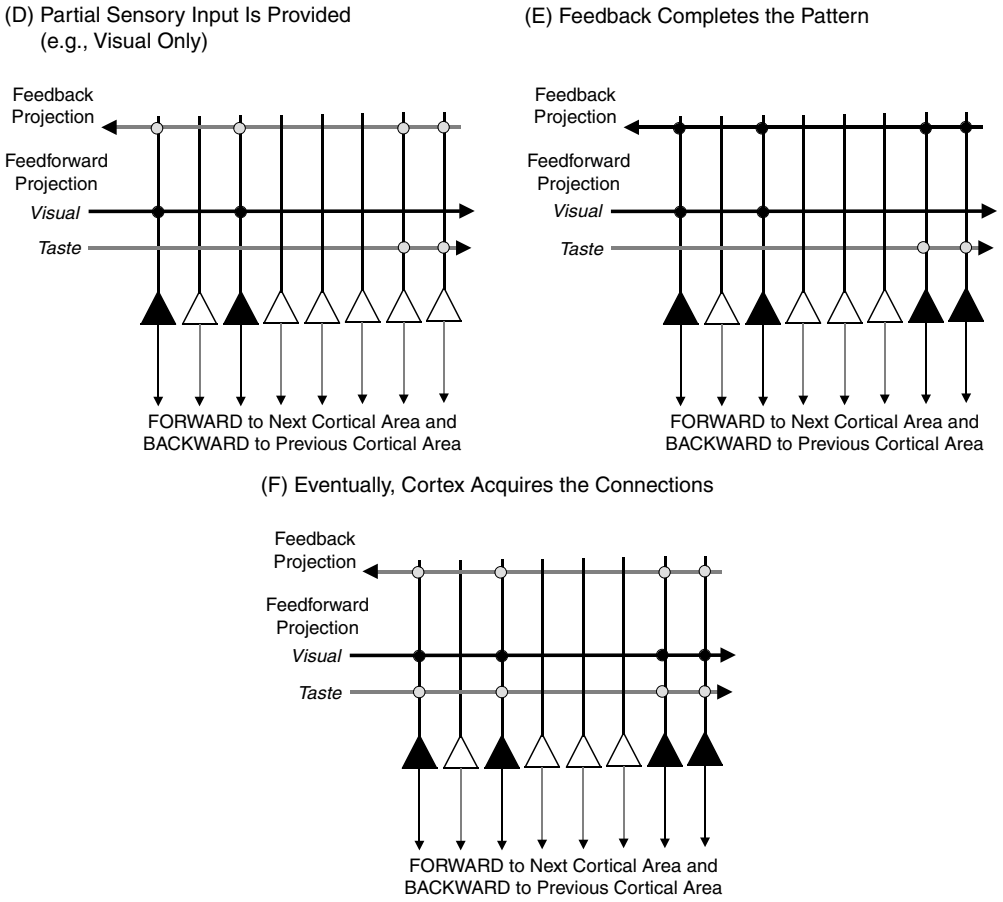


Figure 9.17 ***(continued)*** (D) Later, a partial version of the familiar pattern is presented; a subset of previously trained neurons become active. (E) This partial pattern is eventually transmitted to hippocampus, which completes the pattern. Feedback projections carry the complete representation back to cortex, completing the pattern in cortex. The pattern is recalled. (F) Eventually, local connections between coactive neurons in cortex are strengthened, and cortex can recall entire patterns without the help of hippocampal feedback: The memory is consolidated.

make random weighted synapses on the pyramidal neurons (shown as circles). In general, feedback from the hippocampal region is assumed to sparsify the pattern, as outlined previously in this chapter. Where feedforward and feedback inputs converge (figure 9.17B), neurons are strongly activated. Under the assumption that the cortex performs competitive learning, strongly activated "winning" neurons undergo plasticity—strengthening the

weights from active inputs—while the remaining neurons undergo weight decreases (figure 9.17C). At this point, the pattern and its hippocampal representation are stored.

Consider now what happens when a partial version of a stored pattern is presented to the network: for example, the sight of a stimulus without its taste, as schematized in figure 9.17D. A subset of neurons that were previously trained become active. This partial pattern is projected forward to the hippocampal region and eventually returns as feedback projections (figure 9.17E); the additional feedback input is enough to reconstruct the original stored pattern in cortex. The complete pattern is recalled. Eventually, local recurrent connections between neurons (not shown in the figure) become strengthened, so cortex can recall entire patterns on its own without the help of hippocampal feedback (figure 9.17F); at this point, the memory is consolidated in cortex, and hippocampal-region damage will not disrupt recall.

This basic plan of cortical backprojection has several important implications. First and most important, it provides a plausible interpretation of how sensory representations from the hippocampal region could eventually be adopted in cortex. One problem in theorizing about hippocampal-region function has always been in understanding how very specific activation patterns (e.g., episodic memories) encoded in hippocampus could be transferred to cortical storage. In a computational model, it is easy to hardwire appropriate connections between a hippocampal-region module and a cortical module. However, in the brain, it is less likely that there exists a predetermined blueprint ensuring that, for any neuron X in the cortex that projects to neuron Y in another region, there is a reciprocal connection from Y to X. Worse, even if every neuron in hippocampus did have such reciprocal connections with the rest of the brain, there is huge convergence from the entire brain onto a relatively small number of hippocampal-region cells. How would the hippocampus "know" which of these backprojections were appropriate for the current information? The mechanism of figure 9.17 neatly sidesteps this issue. In Rolls's model, the hippocampus doesn't "know" where to project information; it simply projects everywhere, and storage occurs automatically wherever conjoint feedforward and feedback projections converge.

A second implication of Rolls's theory concerns the mechanism of storage. Assuming that the hippocampus projects to the right place in cortex, how is information stored there? It does not seem likely that, as many computational models propose, the hippocampal output acts as a teaching signal, forcing cortical cells to respond to particular inputs. As we described in previous chapters, such teaching inputs require very specific anatomical

properties that appear to exist in only a few places in the brain (e.g., the mossy fiber connections from dentate gyrus to CA3 and the climbing fiber inputs to cerebellum). The Rolls scenario requires no more than the ubiquitous Hebbianlike learning, in which plasticity occurs to strengthen the connections between any two coactive inputs. Neurobiological mechanisms for Hebbian learning have indeed been observed in superficial layers of cerebral cortex, right where Rolls's model requires that they should be.[39]

One limitation of the system illustrated in figure 9.17 is that it may learn too much: Every time a new input is presented, the hippocampal region will develop a new representation of that input and backproject that representation to cortex for storage. But such constant learning will quickly overload the network, leading to catastrophic interference in which new information overwrites old. Rolls notes that there is a fairly simple way to avoid this problem, and that is to assume that the system does not learn arbitrary information, but only information that is "significant."[40] Other brain areas, such as the amygdala (which is involved in learning the emotional significance of stimuli) and basal forebrain (which is involved in signaling stimulus novelty), could provide this information through projections to cortex. Only in the presence of one or more of these signals would cortical storage be enabled. The next chapter will return to this issue of how subcortical inputs might modulate memory storage for especially significant events.

9.3 RELATIONSHIP TO QUALITATIVE THEORIES: STIMULUS BUFFERING AND CONFIGURATION

Howard Eichenbaum and colleagues Tim Otto, Neal Cohen, and Mike Bunsey have addressed the problem of distinguishing entorhinal hippocampal function on the basis of behavioral and neurophysiological data. They divide the hippocampal region into two basic components: the **parahippocampal region,** which includes entorhinal cortex and nearby perirhinal and parahippocampal cortices (called postrhinal cortex in rats), and the **hippocampal formation,** which includes the hippocampus and the dentate gyrus.

The subiculum appears to be transitional between the two areas. In rats and monkeys, damage to the parahippocampal region devastates the ability to perform tasks such as delayed nonmatch to sample (DNMS) that require maintaining stimulus information over a short interval (e.g., 30–120 seconds). Recordings of neuronal activity in the parahippocampal region during such a task suggest that cells hold firing patterns during the interval.[41] On the basis of these and related data, Eichenbaum and colleagues have

suggested that the parahippocampal region operates as an intermediate store of stimulus information.[42]

One aspect of this storage is the ability to fuse co-occurring stimuli and to allow configuration of stimuli that co-occur with slight temporal displacement (e.g., a lightning flash and a thunderclap). The hippocampal region, by contrast, is hypothesized to play a largely antagonistic role: learning about relationships between individual items. Selective removal of the hippocampus thereby results in a system that tends to overcompress stimulus information at the expense of recognizing individual components. A broader lesion that included the parahippocampal region would eliminate this compression, leaving only the ability to form simpler stimulus-response relationships.

Eichenbaum and Bunsey tested these predictions in a rat odor discrimination paradigm.[43] In these studies, a rat was presented with odor stimuli (e.g., A, B, C, and D) and had to choose whether to respond. The odors were presented in sequential pairs: one odor followed by a second odor. Some odor pairings (e.g., AB+ and CD+) were rewarded; recombinations or **mispairs** of these odors (e.g., AC−, BD−) were not rewarded. Thus, this task required rats to learn about configurations of odors, because any individual odor could be rewarded or nonrewarded, depending on its companion odor.

Figure 9.18A shows the empirical results. Control rats learned to respond to pairs but not mispairs within about 21 training sessions. In contrasts, rats with parahippocampal-region lesion (PRER) did not learn even within 35 sessions.[44] This is consistent with the hypothesis that the parahippocampal region is involved in compressing or configuring stimuli. Most interestingly, rats with selective hippocampal lesions were *facilitated* relative to controls.[45] Once the putatively antagonistic hippocampus was removed, the parahippocampal region was free to work its compression unfettered, and since compression of odor pairs facilitated learning this particular task, learning was speeded.

At the same time, the rats were trained to discriminate odor pairs from **nonpairs,** in which one odor was never associated with reward (e.g., AX−, BY−). To solve this task, the rats needed only to learn never to respond on any trial involving the nonrelational odors X and Y; it was not necessary to attend to stimulus configurations. All that was needed was strong inhibitory weights from X and Y. In this case, the control, PRER, and H-lesion rats all learned at the same speed (figure 9.18B).[46] Again, these results supported the theory: The parahippocampal region (and hippocampus) are needed only for learning involving relationships between cues, not for simple stimulus-response learning.

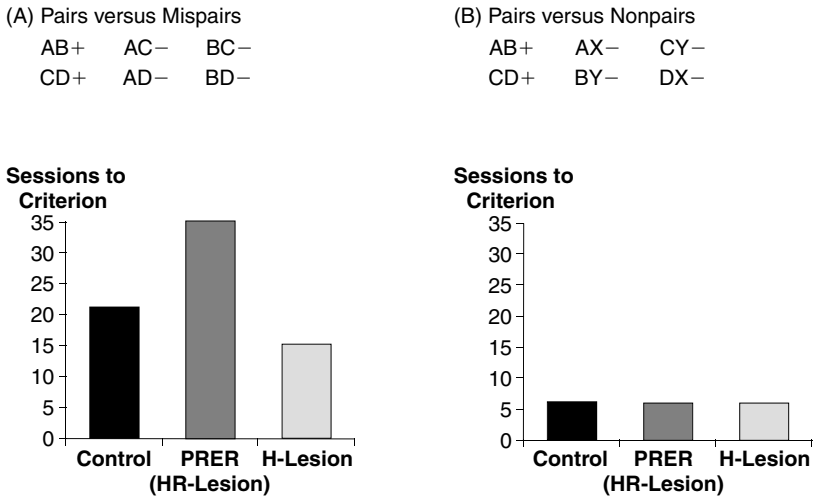


Figure 9.18 The paired associate learning task. (A) One component of the task involved learning to respond to odors when presented in particular pairs (e.g., AB+, CD+) but not when recombined into mispairs (e.g., AC−, BD−). Note that this task cannot be solved by assigning weights to individual components but only by assigning weights to compounds: Thus, A is sometimes positive (when paired with B) and sometimes negative (when paired with C or D). Control rats learned the distinction within about 21 training sessions. Rats with lesions of the perirhinal and entorhinal cortex (PRER), which also disconnect the hippocampus from its primary sensory input and output, were greatly impaired; most did not learn within 35 sessions. This is consistent with the idea that the entorhinal cortex (or, more broadly, the parahippocampal region) compresses the representation of co-occurring stimuli into compound stimuli. By contrast, rats with selective hippocampal lesions (H lesion) were facilitated in learning this discrimination. This is consistent with the idea that the hippocampus normally performs a function antagonistic to the entorhinal cortex; its removal disinhibits entorhinal function. (B) A similar discrimination, with the same odor pairs but with contrasting nonpairs, which include one odor that is not part of a reinforced pair (e.g., X, Y). Note that this task can be solved by attaching strong negative weights to nonpair odors. Control, PRER, and H-lesion rats all learn at equivalent speeds. (Plotted from data presented in Bunsey & Eichenbaum, 1993; Eichenbaum & Bunsey, 1995.)

By now, the reader will have noticed a strong similarity between Eichenbaum and colleagues' account of parahippocampal and hippocampal interplay and that embodied in our own cortico-hippocampal model.[47] Whereas Eichenbaum and colleagues propose that the parahippocampal region configures stimuli, the cortico-hippocampal model assumes that the entorhinal cortex compresses stimulus representations. The chief differences are that Eichenbaum and colleagues include a temporal dimension and that they

consider the entire parahippocampal region, not merely the entorhinal cortex. Otherwise, the two statements of function are quite compatible.

This correspondence is especially encouraging, since these parallel ideas emerged from two very different research traditions: one based on behavioral and neurophysiological observations of rat odor discrimination and the other from computational modeling constrained by anatomical data.

9.4 IMPLICATIONS FOR HUMAN MEMORY AND MEMORY DISORDERS

One of the most pressing reasons for trying to understand hippocampal-region function is that hippocampal-region dysfunction may be an important contributor to the cognitive impairments of Alzheimer's disease. **Alzheimer's disease (AD)** is a progressive, degenerative neurologic illness that may afflict up to four million adults in the United States alone. Currently, it is a leading cause of death among Americans over the age of 60, afflicting an estimated 25% of Americans over age 85. The causes of the disease are unknown, although there appears to be a genetic component, but abnormal elements called **plaques** and **tangles** accumulate in the brain. Neurons begin to degenerate, losing synapses and eventually dying (figure 9.19A). As this occurs, the brain itself begins to shrink or **atrophy** (figure 9.19B), causing behavioral impairments and eventually death.

One of the earliest deficits in AD is memory decline. This is evidenced by failure on such tasks as the **paragraph delayed recall** test, in which the experimenter reads a short story aloud and asks the subject to repeat the story back. The subject gets a point for every item in the story that is repeated correctly (typically, there are about 25 such items). Next, there is a delay of 5–15 minutes, and then the subject is asked to repeat the story once more from memory. A normal young or middle-aged subject may be able to recall most of the paragraph, while a normal elderly subject may recall ten or eleven of the story items. Subjects with AD may do substantially worse. Even a very mild impairment on this kind of test may indicate an increased risk for developing AD.[48] Interestingly, this kind of task—which involves the recall of factual information over a time span of a few minutes—is exactly the kind of memory task that is disrupted in amnesic subjects with medial temporal lobe damage. Therefore, one might suspect that hippocampal damage or dysfunction underlies some of the cognitive deficits in early AD.

As Alzheimer's disease progresses to later stages, the symptoms become devastating, including memory loss, personality changes, loss of initiative (apathy), poor judgment, disorientation of time and place, and depression. Individuals with AD experience a gradual decline in abilities until they

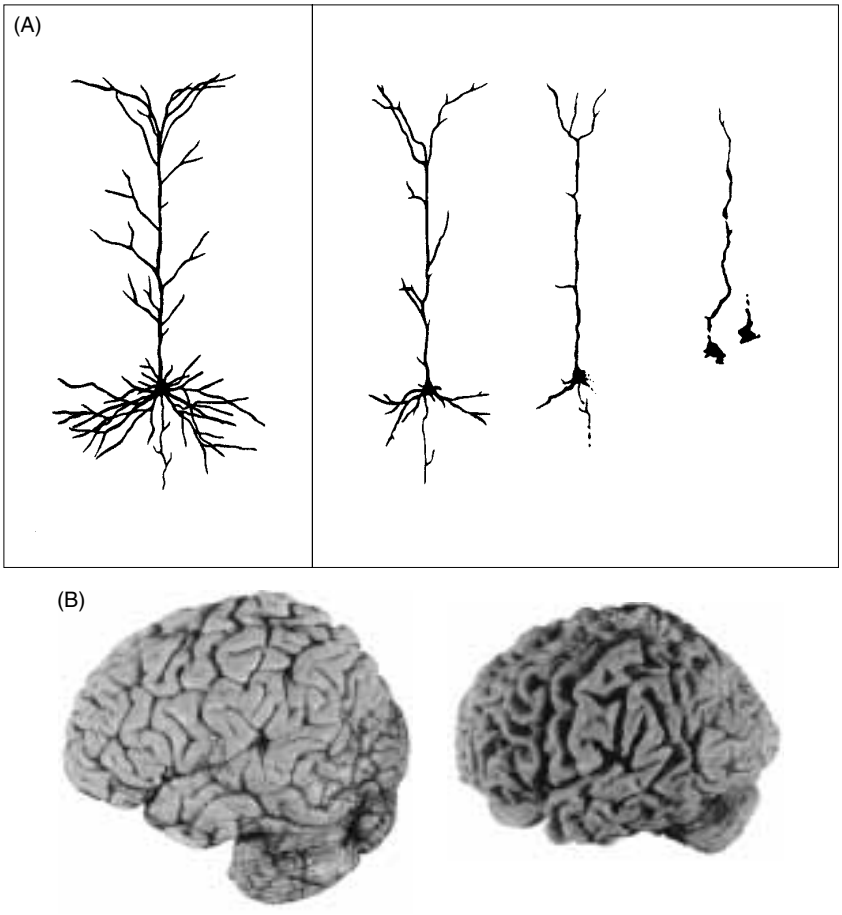


Figure 9.19 (A) Left: A neuron in normal human cortex has processes (dendrites) that branch multiply and widely. Other neurons make synapses on these dendrites, allowing information transfer. Right: Cells from the same area of patients with Alzheimer's disease show progressive degeneration, especially in terms of shrinking of dendritic branches. (Adapted from Kalat, 1995, p. 463, figure 138.) (B) The cerebral cortex of a normal person (left) shows a characteristic wrinkled appearance as the cortex folds in on itself. The cerebral cortex of a patient with Alzheimer's disease (right) is shrunken, with exaggerated spaces between folds of cortex. Atrophy occurs in other brain areas too, including the hippocampal region. (Adapted from Kalat, 1995, p. 462, figure 13.7, which reprinted the pictures courtesy of Dr. Robert Terry.)

eventually require round-the-clock care, including assistance with eating, dressing, bathing, and other daily activities.

Currently, there is no cure for AD. The few medications that are available treat the symptoms. At best, the medicines temporarily arrest the progress of the disease; they cannot reverse the damage or prevent further decline.

Because of this, it is vital to identify individuals who are at risk to develop AD before they begin to show cognitive decline and memory impairments. Memories, once lost, are gone forever.

Recent research has produced findings that may allow early detection of which individuals are most at risk to develop AD in the future. In some elderly individuals, the hippocampus and entorhinal cortex show atrophy while other nearby brain structures appear intact. This can be seen by using neuroradiography such as **magnetic resonance imaging (MRI).** These techniques provide a picture of a slice through the brain (or body) without harming the individual being imaged. Figure 9.20 shows two MRIs of human brains, taken in the horizontal plane, meaning that they would parallel the floor if the individual were standing. Figure 9.20A shows a normal brain; the two hippocampi are in the boxed areas. Figure 9.20B shows an individual with hippocampal atrophy; the hippocampi are considerably shrunken.[49] There is little visible shrinkage of other nearby areas; the atrophy appears to be largely confined to the hippocampus itself.

The kind of hippocampal atrophy that is shown in figure 9.20B has been observed both in patients diagnosed with mild AD and in elderly individuals who have cognitive impairments that are suggestive of possible AD.[50] Thus, hippocampal atrophy may be an indicator of which individuals are most at risk to subsequently develop AD.[51] Not every individual with

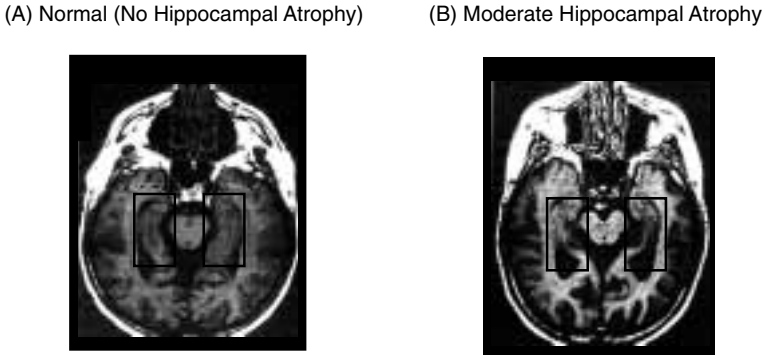


Figure 9.20 Two magnetic resonance images (MRIs) of the human brain, showing slices through the brain that would be parallel to the floor if the subject was standing upright. The nose is at the top of the images, and the back of the head is at the bottom. The hippocampi lie in the areas outlined in black. (A) Normal (nonatrophied) hippocampi. (B) Moderately atrophied hippocampi. The kind of atrophy shown in B could indicate that the individual is at risk for future development of the kind of cognitive impairments that are associated with AD (Golomb et al., 1993). (Images adapted from de Leon et al., 1993a.)

hippocampal atrophy will go on to develop AD, but hippocampal atrophy may be one warning sign; individuals with atrophy could then be monitored closely to see whether there are further signs of decline, and these individuals could be started on drug therapies as early as is appropriate.

If hippocampal atrophy is indeed a predictor of cognitive decline and AD, then it ought to be possible to estimate hippocampal atrophy by behavior alone: If an individual starts to perform poorly on hippocampal-dependent tests, then that might indicate that the individual is experiencing hippocampal atrophy. One example is the paragraph recall task described above; poor performance on this test is correlated with hippocampal atrophy.[52] However, this task is sensitive to many different kinds of brain dysfunction that result in memory impairments.

What is really needed is a task that is specifically designed to tap into hippocampal-region function. Then, poor performance on this task would be a much better indicator that hippocampal atrophy—rather than some other kind of damage or disease—was present.

Behavioral Measures of Hippocampal Atrophy

The cortico-hippocampal model suggests that selective hippocampal damage might lead to overcompression deficits: a tendency to perceive co-occurring stimuli (or stimulus features) as compound objects, leading to a difficulty in subsequent discrimination. Further, the previous chapter noted that broad HR lesion that includes the entorhinal cortex might have a similar effect, at least for tasks involving compression within a single modality that might be mediated by unimodal sensory cortices. The implication is that hippocampal damage (with or without conjoint entorhinal damage) might lead to an overcompression deficit. Given training with co-occurring stimuli, lesioned subjects will tend to overcompress and therefore perform poorly when familiar stimuli are presented in novel recombinations. The same logic will hold true for stimulus features. Remember that the cortical network has no *a priori* way of knowing whether two co-occurring inputs are different stimuli or different features of the same stimulus.

For example, consider the visual discrimination task shown in figure 9.21, which was developed in our laboratory as a test of hippocampal-region function.[53] There are eight pairs of objects (four are shown in figure 9.21A), and each object has the features of color and shape. (In figure 9.21, colors are approximated as gray levels.) Within each discrimination, only one feature varies across the object pair: The first two discriminations differ in color but not shape, and the last two differ in shape but not color. Subjects in our study

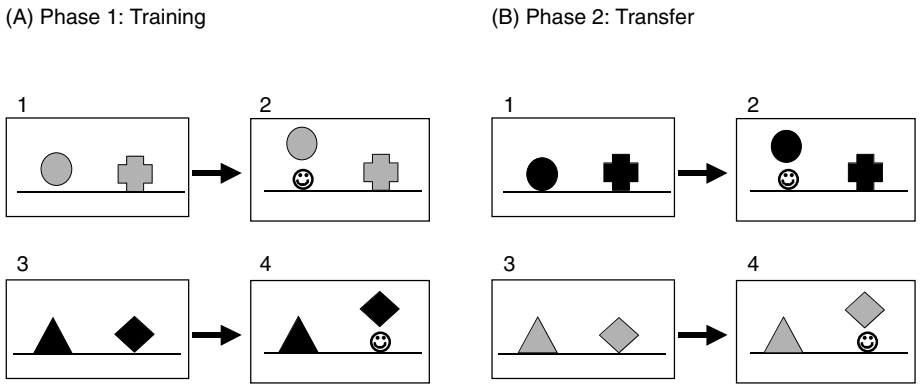


Figure 9.21 A color-shape discrimination task from Myers, Kluger, et al. (1998). (A) Phase 1 (training phase) consists of eight pairs of colored shapes (here, colors are approximated by gray levels). For example, one pair might consist of a same-colored circle and cross (1). The subject saw the two objects on the computer screen (1) and chooses one. The chosen object is raised (2), and if the subject's choice was correct, a smiley face is revealed underneath. Here, the circle is always the correct choice. Another pair might consist of a same-colored triangle and diamond (3), with the diamond always correct (4). For both these discrimination pairs, the shape but not the color is relevant with respect to predicting the smiley face's location. Other discrimination pairs differ in color but not shape (e.g., a red spiral versus a yellow spiral). Phase 1 continues until the subject reaches criterion performance, correctly responding on all eight of discrimination pairs. (B) Phase 2 is a transfer phase. Here, objects are recombinations of familiar colors and shapes, maintaining the relevant features from phase 1. Thus, given the circle-cross discrimination pair (1), the circle is still the correct choice (2). Given the triangle-diamond discrimination pair (2), the diamond is still the correct choice. Color is still irrelevant in these discriminations. Subjects who learn to pay attention only to the relevant features in phase 1 (e.g., choose circle over cross, regardless of color) should perform close to perfectly in phase 2. By contrast, subjects who learn responses based on all (relevant and irrelevant) features of the objects in phase 1 (e.g., choose gray circle over gray cross) should be at a loss when new combinations (e.g., black circle versus black cross) are presented in phase 2.

are shown discrimination pairs, one at a time over many trials, and must learn to choose the rewarded object of each pair, regardless of the objects' left-right position. Eventually, normal subjects come to give the correct response on each trial.

There are at least two strategies that a subject may adopt to solve this kind of task. The first strategy is to notice that, within each discrimination, one stimulus feature is relevant and one is irrelevant. Thus, in discrimination 1 of figure 9.21A, color is relevant and shape is irrelevant. A rule that would guide correct responding is: Choose the red object over the yellow object, regardless of shape. Similar rules can govern responding in the other discriminations.

Note that it is not necessary that subjects explicitly form a verbal rule; a simple associative learning system (such as the Rescorla–Wagner model) will also learn to weight relevant inputs heavily and to ignore irrelevant inputs.

An alternative strategy for solving this discrimination task would be to treat all stimulus features equally and compress them into compound percepts. In effect, subjects might learn to choose the red-square over the yellow-square. This kind of solution requires the ability to form compressed or configured representations of stimulus inputs.

Either strategy can yield perfectly accurate performance on the original discriminations. However, in our studies, we included a second, transfer phase as schematized in figure 9.21B. Here, the familiar stimulus features are recombined into novel objects. The recombination is done according to strict rules: Features that were relevant in phase 1 have the same meaning in phase 2, but now they are paired with different irrelevant features. Thus, whereas the red square but not the yellow square predicted reinforcement in phase 1, the red arrow but not the yellow arrow predicts reinforcement in phase 2.

Now the strategy used to solve phase 1 makes a critical difference. If a subject had learned to ignore the irrelevant features in phase 1, phase 2 performance should be close to perfect. The original rule to choose red over yellow (regardless of shape) in phase 1 still yields the correct answer in phase 2. However, if a subject had learned about stimulus configurations in phase 1, phase 2 performance should be close to chance. The original rule to choose a red-square over a yellow-square is of little use when one is confronted with a red-arrow and a yellow-arrow—two new objects. To master phase 2, the subject would have to learn an entirely new set of rules, and phase 2 learning might take fully as long as the original phase 1 learning.

On the basis of the predictions of our cortico-hippocampal model, we conjectured that normal subjects, with functioning hippocampal regions, would tend to favor the more general solution: Hippocampal-mediated representations would differentiate predictive cues (such as the color in discrimination 1) and de-emphasize or ignore redundant ones (such as the shape in discrimination 1).

Moreover, we expected that irrelevant features should be largely ignored in phase 1 learning, leading to very good phase 2 transfer performance. In contrast, our cortico-hippocampal model predicts that hippocampal-region damage should impair this differentiation process, while cortical compression would proceed unchecked. Thus, we expected that subjects with hippocampal damage might tend to respond to stimulus compounds in phase 1, leading to poor phase 2 performance.

To test our hypothesis, we tested a group of twenty elderly individuals on this task, recruited through an aging study at New York University. All had been given thorough psychological testing as well as neuroimaging like that in figure 9.20. All the individuals were judged cognitively normal on the basis of their performance on the psychological tests, but twelve individuals showed signs of very mild atrophy to one of their hippocampi, much like that seen in figure 9.20B.

Figure 9.22A shows that there was no significant difference between the atrophy and no-atrophy subjects in terms of performance on psychological tests such as the paragraph recall test. Therefore, the subjects with atrophy were not yet showing cognitive decline on standard measures of memory.

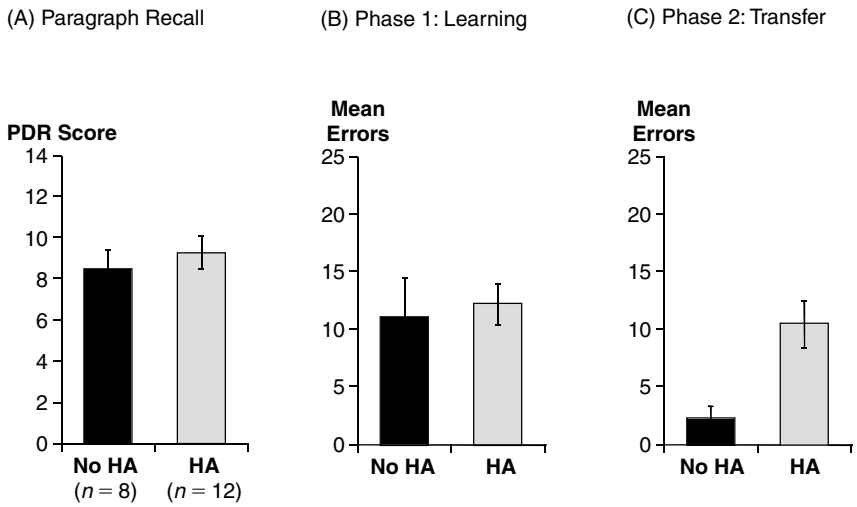


Figure 9.22 (A) Performance on the paragraph delayed recall (PDR) has previously been shown to correlate with (and therefore predict) hippocampal atrophy. In the current study, however, PDR did not discriminate between subjects without hippocampal atrophy (noHA) and those with mild hippocampal atrophy (HA). All subjects averaged recall of about eight to ten items from the paragraph after a 5- to 10-minute delay; this performance is about normal for elderly subjects. (B) Performance on phase 1 of the colored shape discrimination, in terms of total errors, is not significantly different between the subjects with no hippocampal atrophy (noHA) and with mild hippocampal atrophy (HA). (C) However, there is a dramatic difference on performance in phase 2: Subjects with no atrophy performed close to perfect, as expected, while subjects with mild atrophy performed significantly worse. Thus, this simple transfer task may discriminate between subjects with and without hippocampal atrophy. (From data presented in Myers, Kluger, et al., in prep.)

When tested on the visual discrimination task, all twenty subjects were able to master the first phase of learning with the original eight discriminations with the colored shapes. Figure 9.22B shows that there was no significant difference between atrophy and no-atrophy subjects in terms of total errors to reach criterion performance on this phase.

On the other hand, there was a dramatic difference on the transfer phase (figure 9.22C).[54] Subjects with no hippocampal atrophy performed close to perfect on this phase, averaging only about two errors. Subjects with hippocampal atrophy performed very differently. These subjects averaged about twelve errors—approximately as many as in phase 1. This suggests that these subjects had to learn phase 2 from scratch—The prior experience in phase 1 was no help at all. This in turn suggests that the subjects with hippocampal atrophy may have formed overcompressed rules in phase 1 that did not apply to the recombined features in phase 2.

Tasks such as our transfer test described above may represent a possible way to identify which elderly individuals may have hippocampal atrophy before they begin to exhibit serious cognitive decline. This classification is made on the basis of abnormal behavior on a task that may reflect hippocampal dysfunction. Our hope is that this and similar tasks may eventually be used in general practice to evaluate nondemented elderly individuals. Elderly individuals could receive this kind of simple test as a part of their annual checkup. When and if an individual's performance begins to show signs of hippocampal dysfunction, this would signal that the individual should undergo more elaborate testing, including MRI analysis and other means to determine risk for AD.

At present, though, the results described above represent only the first step in developing such a diagnostic test. Only a small sample of elderly subjects have been tested, and among these, even those described as having hippocampal atrophy have only very mild degrees of atrophy. We do not yet know how individuals with more pronounced atrophy may perform. More important, even if the task is reliable in detecting hippocampal atrophy, we do not yet know whether it is reliable in predicting risk for AD. The subjects here might indeed have hippocampal atrophy, but as a result of some other disease or prior trauma. For instance, some kinds of stroke can result in hippocampal atrophy; in that case, the presence of atrophy would not necessarily mean that the individual was at risk to develop AD. For now, one important project in our laboratory is to track the individuals who participated in our original study and find out which of them do go on to show cognitive decline and eventually develop AD. Only then can we fully evaluate the clinical potential of this approach.

Entorhinal Versus Hippocampal Atrophy in AD

A final important issue concerns the relative contributions of the hippocampus and entorhinal cortex. So far, we have described only atrophy of the hippocampus as an early indicator of AD. However, the entorhinal cortex also shows signs of pathology early in AD—perhaps even before hippocampal atrophy becomes apparent.[55] Entorhinal atrophy is harder to evaluate than hippocampal atrophy, for anatomical reasons.

As is evident in figure 9.20, the hippocampus is visibly different from its surrounding structures. It is possible to precisely delineate the hippocampus's boundaries for the purposes of calculating its volume and estimating shrinkage. The entorhinal cortex, by contrast, does not have a precise boundary. It is difficult to tell where it ends and the next cortical area begins. Many researchers are currently working to develop procedures for reliably measuring the precise volume of the entorhinal cortex.

SUMMARY

- The entorhinal cortex is a periallocortical structure within the hippocampal region, meaning that its form is intermediate between six-layered neocortex and two-layered allocortex. It receives highly processed, multimodal sensory input and projects to the hippocampus; in turn, hippocampal outputs project to entorhinal cortex and from there back to the cortical areas where they arose.
- On some tasks, memory impairment simply grows as a function of how much hippocampal-region tissue is lesioned. But more recent studies have suggested qualitative differences in impairment based on precise lesion extent, suggesting that a lesion that is precisely limited to hippocampus may spare some entorhinal function.
- A simple entorhinal model may be constructed similar to the piriform cortex model: Clusters of nodes complete to respond to inputs; representations of similar and co-occurring (redundant) inputs are compressed.
- The cortico-hippocampal model suggests that this entorhinal compression may survive selective hippocampal damage, which disables a normally competing hippocampal differentiation function. The result is that selective hippocampal lesion results in a tendency to overcompress information, and this is consistent with existing data.
- Other computational models have suggested that the entorhinal cortex is involved in stimulus configuration and the backprojections from hippocampus

to entorhinal cortex and beyond provide a possible substrate for memory consolidation.

• Eichenbaum and colleagues have suggested that the entorhinal cortex is an intermediate-term buffer, one aspect of this buffer is the ability to configure representations of items that occur with slight temporal displacement. This proposal is quite consonant with the cortico-hippocampal model's account.

• A discrimination task, based on the model predictions and animal data, shows some promise for detecting the very mild hippocampal atrophy that may be an indicator of risk for subsequent development of Alzheimer's disease.

APPENDIX 9.1 SIMULATION DETAILS

The H-lesion model of figure 9.7C, including entorhinal network, was originally presented in Myers, Gluck, and Granger (1995).

Briefly, the entorhinal network consists of a single layer of 100 nodes, divided into five nonoverlapping patches of 20 nodes each. Nodes in one patch are all reciprocally connected with a single local inhibitory feedback cell. Each node receives weighted connections from all external inputs, and the weights are initialized from a uniform distribution and normalized so that the sum of all weights to one node equals 1.0. On each trial, each node n computes activation V_n as a weighted sum across all external inputs i:

$$V_n = \sum_i w_{in} y_i$$

In each patch, the node with maximal activation is the winner. The output of the winning node w is set to $y_w = 1.0$, while the local inhibitory cell silences all other nodes in the patch (output $y_j = 0.0$ for all $j \neq w$). The winning nodes (one per patch) update their weights as

$$\Delta w_{iw} = \alpha I_i (y_w - V_w)$$

where $\alpha = 0.001$ for winning nodes and $\alpha = 0.0001$ for all other nodes.

In the H-lesion model of figure 9.7C, the entorhinal network output provides the training signal for the internal-layer nodes of the cortico/cerebellar network. For each internal-layer node in the cortico/cerebellar network, the desired output is a weighted sum of the activations of the nodes in the entorhinal network. These weights are initialized from a normal distribution and fixed thereafter. Otherwise, the cortico/cerebellar network is similar to that found in previous chapters. (Full details of the entorhinal network and H-lesion model are given in the appendix to Myers et al., 1995.)