

This excerpt from

Gateway to Memory.

Mark A. Gluck and Catherine E. Myers.

© 2000 The MIT Press.

is provided in screen-viewable form for personal use only by members of MIT CogNet.

Unauthorized use or dissemination of this information is expressly forbidden.

If you have any questions about this material, please contact
cognetadmin@cognet.mit.edu.

8 Stimulus Representation in Cortex

The computational models of hippocampus discussed in previous chapters have considered the role of the hippocampal region in stimulus representation and memory. However, most of these models acknowledge that the hippocampus does not operate as an isolated structure. Any information processed by the hippocampus must eventually be transferred to other brain areas for long-term storage. Models of hippocampal consolidation make this point explicitly, but our cortico-hippocampal model¹ also assumes that representations developed in the hippocampal region are eventually adopted by other brain areas such as the cortex and cerebellum.

Most of the models that we have discussed so far assume that the hippocampal region operates directly on sensory input (such as CSs). This is also an oversimplification. The hippocampal region is the culmination of a long and intricate processing chain. Sensory inputs from receptors such as retinal transducers and taste buds travel through the thalamus and into areas of cerebral cortex that are specifically devoted to processing different kinds of sensory information. From there, information travels to higher cortical areas, which combine and integrate across sensory modalities, before finally reaching the hippocampal region.

Thus, among its other functions, cerebral cortex preprocesses hippocampal-region inputs. To fully understand what the hippocampal region is doing, it is therefore necessary to have some understanding of its inputs—and hence what the cerebral cortex is doing.

In this chapter, we review some basics of cortical representation. We show how certain types of network models can be related to cortical architecture and physiology. We then present a specific model that combines a cortical module with a hippocampal-region module; with this model, we then explore how these brain systems might interact. Finally, we present an example of how research into cortical representation is leading to a real-world application to help children who are language learning-impaired.

8.1 CORTICAL REPRESENTATION AND PLASTICITY

The cerebral cortex is a grayish sheet covering most of the mammalian brain and containing the cell bodies of neurons. The output processes of these cells, **axons**, form the underlying “white matter” that makes up much of the bulk of the brain in higher species. Most mammalian cortex is only a few millimeters thick but has six distinct layers, each characterized by the presence or absence of various cell types (figure 8.1A). Only mammals have six-layered cortex; reptiles and birds have a simpler kind of cortex—sometimes called **paleocortex** or **allocortex**—with only two layers (figure 8.1B). Six-layer cortex is often called **neocortex** or **isocortex**, reflecting the assumption that it developed later in evolution. However, there are a few places in the mammalian brain that are also two-layered paleocortex, and these may be evolutionary remnants from which neocortex developed.

Although overall brain weight is related to an animal’s body size, the cerebral cortex is disproportionately elaborated in higher species. For example, if a rat’s brain were enlarged to be the same size as a human brain, many parts would be similar in the two species (figure 8.2). However, the rat’s olfactory

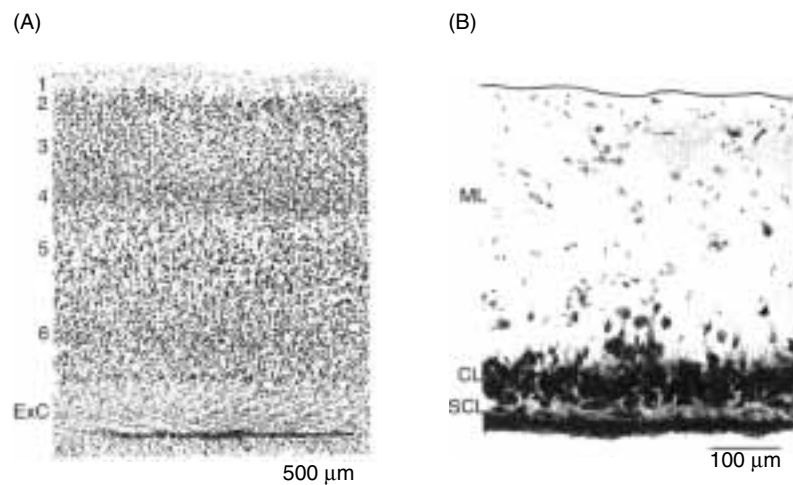


Figure 8.1 (A) High-power photomicrograph of a slice through mammalian neocortex (rat), showing complex cellular and laminar structure. Cortical layers 1 through 6 are labeled. The top of the figure corresponds to the outer layer of cortex; ExC = external capsule lies below the cortex. The scale bar near the bottom of the picture equals 500 μm . (B) High-power photomicrograph of a slice through reptilian cortex (turtle), showing much simpler two-layered structure. The ridge near the top of the picture is the outer layer of the brain. CL = cellular layer; ML = molecular layer; SCL = subcellular layer (sometimes considered a third layer). The scale bar at the bottom of the picture equals 100 μm . (Reprinted from Zigmond et al., 1999, p. 1289, figure 50.3C,E.)

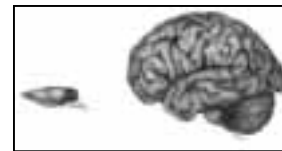
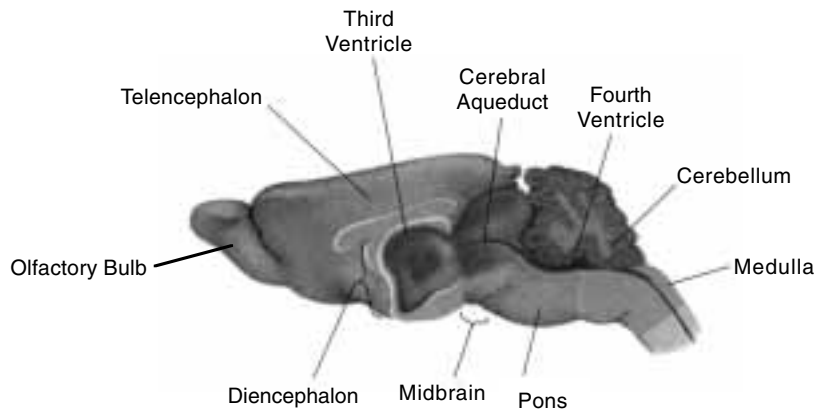
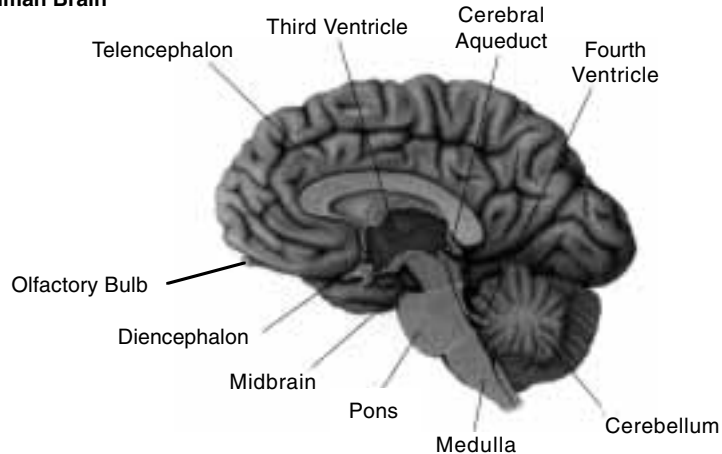
Rat Brain**Human Brain**

Figure 8.2 Mammalian brains show a variety in size and complexity but also share many features, including the presence of six-layered neocortex. For example, the rat brain (top) and human brain (bottom), drawn to be equivalent in size, show many of the same structures (a few are labeled for comparison). However, different structures are elaborated in different species. The rat's olfactory bulb is disproportionately large relative to the human's, presumably reflecting the importance of olfactory stimuli to the rat. Conversely, neocortex reaches its highest elaboration in humans, resulting in the characteristic wrinkled surface, which allows the large cortical sheet to fit in the confines of the skull. (Adapted from Bear et al., 1996, p. 176, figure 7.21.)

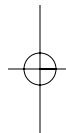
system would be proportionately larger than the human's, while the rat's neocortex would still only be a fraction the size of the human's. The cortex of an adult human would measure about 1.5 square feet if spread out into a flat sheet; the characteristic wrinkled appearance of the human brain reflects the fact that the cortex has to fold to fit inside the confines of the skull. The rat cortex, by contrast, fits quite comfortably inside the rat skull without wrinkles.

Most areas of neocortex share the same basic organizational structure: Tissue from different areas of neocortex looks very similar to that shown in figure 8.1A. However, different areas of neocortex process different kinds of information, and this differentiation arises from the different inputs they receive.²

For each sensory modality, the first cortical processing occurs in a specific region: **primary visual cortex** (or **V1**) for vision, **primary auditory cortex** (or **A1**) for sounds, and so on. From there, sensory information progresses to higher sensory areas, which integrate information further, first within and then across modalities. One of the most interesting features of primary sensory cortex is that it is organized **topographically**. That is, *each region of cortex responds preferentially to a particular type of stimulus, and neighboring cortical regions respond to similar stimuli. Thus, it is possible to draw a "map" of sensory space on the cortical surface.*

For example, **primary sensory cortex** (or **S1**) is a thin strip of cortex running down each side of the brain in humans (figure 8.3A). By recording the activity of individual neurons in S1 while touching various areas of the skin, it was found that different neurons respond maximally when different body regions are stimulated. Thus, some cells respond to touch stimulation on a particular finger, some on a region of the face, and so on. If this recording is done for a large number of S1 neurons, it is possible to draw a "map" of the body on S1 so that each body part lies over the cortical region that responds to it (figure 8.3A).

Moreover, adjoining areas of S1 contain cells that tend to respond to adjoining areas of body surface. Parts of the body that are especially sensitive to touch, such as fingers and lips, activate larger areas of S1. The result is a distorted neural representation of the human figure, often called a **homunculus**, with exaggerated hands and lips but a greatly shrunken torso. Primary somatosensory cortex in animals shows similar organization except that the homunculus is replaced by a distorted figure of the species in question, and this figure will be distorted to reflect body areas that are more or less critical for that animal. Thus, primates receive a great deal of touch information through their fingers, so these are disproportionately elaborated



across modalities and across mammalian species. The output from primary sensory cortex travels to higher-order cortical areas, where it appears that the obvious topographic ordering may be lost, or else information may simply be coded in a topographic manner that researchers don't yet fully understand.

For many years it was believed that cortical topographic maps were fixed—predominantly hard-wired from birth and subject to fine-tuning only during a **critical period** that extended through the first few months of life. Certainly, the maps are especially plastic at that early period. For instance, in a normal cat, primary visual cortex (V1) contains alternating patches of cells that respond to input from either the left eye or the right eye. If a kitten is reared from birth with its left eye sealed shut, V1 will reorganize accordingly: Proportionally more area will respond to input from the active right eye than from the inactive left eye. Another way of saying this is that the representation of the active eye expands, and the representation of the inactive eye is compressed. If the kitten matures into a cat and the sealed eye is then opened, the visual map in V1 cannot reorganize enough to accommodate the drastic change. The result is that the visual cortex cannot process information from the left eye; the animal will be functionally blind in the left eye, even though that left eye is now fully operational.³

However, more recent experimental findings have demonstrated that cortical plasticity can and does occur in adult animals. For example, when a limb is amputated, the part of S1 representing the lost limb is deafferented, meaning that it receives no more sensory input. Rather than allowing that region of cortex to remain idle, nearby areas on the homunculus may “spread” into the vacated space.⁴ If two or more fingers are always stimulated together, so that the different fingers always experience exactly the same sensory inputs, the parts of S1 representing those fingers may fuse, so that a single area comes to represent the input from both fingers.⁵

One of the most important findings in this area of research in recent years is that cortical reorganization can occur as a function of learning. In one study, Michael Merzenich and colleagues first mapped out the topographic representations in the area of S1 of the monkey that encodes stimulation to the hand (figure 8.4A).⁶ This was done by inserting recording electrodes into various sites in S1 (figure 8.4C) and then stimulating various regions of the hand (figure 8.4B) to see how skin surface mapped to cortical area. The resulting map, shown in figure 8.4D, revealed a neat topography, in which different strips of cortex responded to stimulation of each of the five fingers and within each strip there was an orderly progression from distal sites (finger-tips) to proximal sites (near the base of the finger).

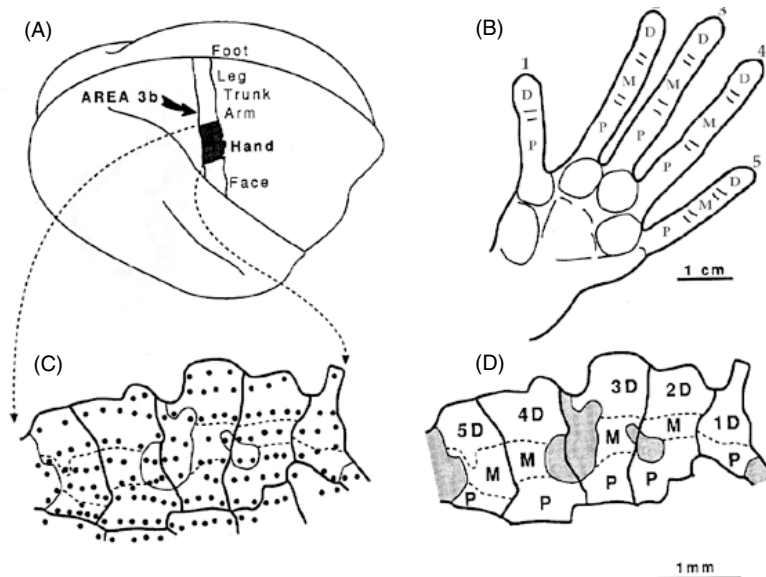
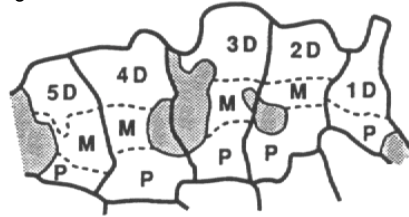


Figure 8.4 (A) The somatosensory cortex of the owl monkey brain contains a small area containing a topographic representation of the hand. (B) The palm of the monkey hand can be divided into areas according to finger (1–5) and site (P = proximal, M = middle, D = distal). (C) Recording electrodes are lowered into S1 at various sites (dots) to record neural activity while different areas of the hand are stimulated. (D) The recordings reveal a topographic map: The five fingers are represented in roughly parallel strips (shown here running from left to right). Within each strip, there is an orderly progression from distal to proximal sites. Dark patches responded to stimulation on the back (hairy side) of the associated finger. The boundary lines in C are drawn to match those discovered in D. (Adapted from Jenkins et al., 1990, figure 1.)

Next, Merzenich and colleagues trained the monkeys for several weeks on a task that required touching a textured spinning disk with the tips of fingers 2 and 3. During training, these fingertips received extensive stimulation in a meaningful context, since the stimulation was associated with obtaining reward. After several weeks, the areas of S1 representing these two fingertips expanded (figure 8.5)—sometimes by a factor of 3.⁷ The representations of the other, unstimulated fingers shifted and even shrank to make room for the expanded fingertip representations.

In another study, Merzenich and colleagues considered changes to primary motor cortex (M1), a strip of cerebral cortex that lies just forward of primary sensory cortex (S1). Like S1, M1 is laid out in the form of a homunculus with exaggerated lips and hands. Stimulation of a region in M1 can result in motor movements in the corresponding part of the body. When monkeys

(A) Before training



(B) After training

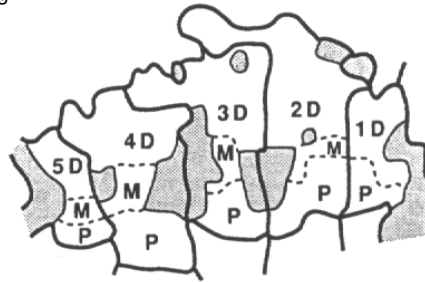


Figure 8.5 Cortical map reorganization following behavioral training (Jenkins et al., 1990). (A) Hand representation within somatosensory cortex S1 of a normal adult owl monkey shows a regular topographic layout, with two rough axes encoding finger (1–5) and position (P = proximal, M = medial, D = distal). The fields encoding distal portions of each finger are approximately equal in size. (B) After several weeks of training on a behavioral task that involved fine sensory stimulation of the tips of fingers 2 and 3 (and occasionally 4), the map changes. The area encoding the stimulated skin surfaces (2D and 3D) are expanded relative to the other fingers and relative to their own original size (A). The representation of 2D has expanded from 0.32 mm² to 0.92 mm²—nearly a threefold increase; the representations of 3D and 4D increase by about 33% and 25%, respectively. The areas encoding nonstimulated areas have shifted to accommodate these expansions, and the area encoding the nonstimulated finger 5 has shrunk to about 80% of its original size. (Adapted from Jenkins et al., 1990, figure 3B,D.)

were trained on a task requiring skilled use of digits, the M1 representation of those digits expanded, while the amount of space devoted to representation of other, nontrained areas (such as wrist and forearm) shrank.⁸

These and related studies suggest an important principle of cortical plasticity: *The representations of stimuli compete for space in the cortical map.* Stimuli that are important, such as those required to execute a task or those that predict reward, increase their areas of cortical representation, and this expansion occurs at the expense of other, nonpredictive stimuli. The topographic nature of the map in primary sensory cortices is generally preserved during these changes, as is shown in figure 8.5B.

The idea that cortical representations are differentiated and compressed on the basis of the predictive nature of stimuli and whether they co-occur is clearly reminiscent of the representational ideas discussed in chapters 6 and 7 in the context of hippocampal-region function and our own cortico-hippocampal model. One important unresolved question is: To what degree do these representational changes in cortex reflect hippocampal-mediated learning, or is this kind of cortical plasticity—observed by Merzenich and others through electrophysiological studies—independent of hippocampus? The answer awaits further experimental studies.

Computational models of cortex have been developed that are consistent with features of cortical anatomy and physiology and that can be used to explore cortical plasticity and the development of topographic maps.⁹ In the next section, we first describe the general principles of these models (often called **competitive** or **self-organizing** networks) and then show how they can be applied to model specific regions of cortex. Next, we show how such a cortical model can be integrated with our existing model of hippocampal-region function, to examine how the two systems might interact to adapt stimulus representations during learning.

8.2 COMPUTATIONAL MODELS

Competitive Learning and Topographic Maps

The error-correcting networks that were described in earlier chapters are **supervised** learning systems, meaning that they learn on the basis of the error between network response and some target. Supervised learning works well in many domains in which the desired response is known or can be inferred. For example, in classical conditioning, the response (such as an eye-blink) should predict reinforcement (such as an airpuff US). A network that performs error correction on the basis of the difference between its response and the US is performing supervised learning.

An **unsupervised** learning system, however, does not learn to produce any predefined desired output. Instead, the network discovers underlying regularities—statistically salient features—of the inputs it is given. For this reason, these networks are also called self-organizing networks, since they do not depend on external reinforcement (such as a US). The autoassociative network discussed in chapter 5 was one kind of unsupervised network. Another class of unsupervised network is a competitive network, meaning that the nodes in the network compete with one another for the right to respond to stimuli. Competitive networks (or networks that embed competitive learning principles) have been considered by a variety of researchers.¹⁰

An example of a competitive network is shown in figure 8.6A. It consists of a single layer of nodes, each of which has weighted connections with every input element. The nodes are organized into nonoverlapping clusters. Within a cluster, only one node can become active at a time. Thus, nodes within a cluster compete with one another for the right to respond. The winner will be the node that is most strongly activated by the current input pattern; it will become maximally active and will inhibit the activity of the other nodes in the cluster. For example, the input pattern in figure 8.6B activates one node in

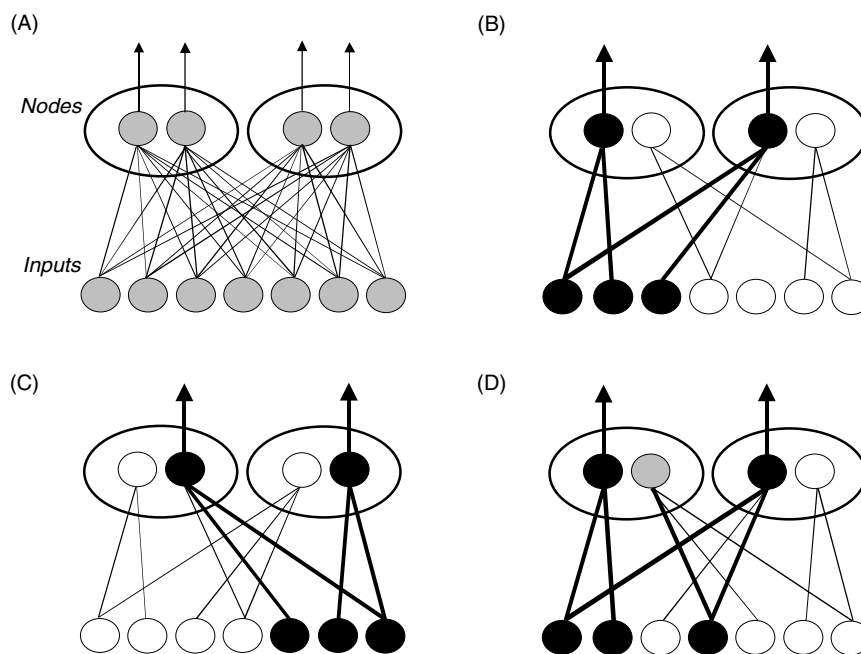


Figure 8.6 (A) Schematic of a competitive network, including seven inputs and four nodes organized into two two-node clusters. Nodes within a cluster compete with each other, meaning that only one can be active to a particular input. The winning node responds to the input; other nodes in the cluster are silenced. The winning node also undergoes plasticity to make it more likely to respond to similar inputs in the future. Over many training trials, the network discovers underlying regularities in the input patterns, so similar inputs generate similar activity patterns over the nodes. (B) Example of a network that has been trained on two input patterns. Here, one input pattern activates one node in the left cluster and one node in the right cluster; other nodes are suppressed. (C) A different input pattern generates a very different activity pattern. (D) A novel input pattern, which shares some features with the pattern in (B), generates an intermediate response. In the left cluster, only one node is activated; it wins the competition by default. In the right cluster, two nodes are partially activated; the stronger one wins and suppresses the other. The result is that the network will respond to this novel input in the same way as it responded to the input in (B); hence, the network will classify the novel input as belonging to the same category as the input in (B).

each cluster; the other nodes are silenced. Within each cluster, the weights from active inputs to the winning node are strengthened to make that node even more likely to win the competition next time the same input pattern is presented. A very different input pattern may generate a very different pattern of activity over the clusters (figure 8.6C).

If there are N nodes in a cluster, each node in the cluster can be considered to represent one of N categories. Each input pattern will be categorized depending on which node in the cluster is activated. For example, if input stimuli range in color along a red-green axis, then one node in a two-node cluster might respond maximally to red stimuli while the other might respond maximally to green stimuli; the color of an arbitrary input can be classified as “red” or “green” by observing which node in the cluster responds to it. Thus, without any external teaching signal, the network learns to classify the inputs. If a novel input is shown that shares some features with one of the trained input patterns (figure 8.6D), it will tend to activate the nodes that are activated by similar inputs. Thus, the novel input will be classified according to the same rules as the trained inputs.

This simple idea can be elaborated in many ways. One possibility is a network with a single, large cluster of nodes arranged in a two-dimensional array (figure 8.7). Each node receives weighted connections from all the inputs; initially, these weights are random. When an input is presented, one of the nodes is maximally activated and wins the competition (e.g., node X in the figure). The weights from active inputs to X are then increased a bit, so that next time the same input is presented, X is even more likely to win the competition again. Then, the immediate neighbors of X (e.g., nodes labeled Y in the figure) undergo similar but smaller weight increases. More distant neighbors (e.g., nodes labeled Z in the figure) may have increasingly smaller weight increases. MathBox 8.1 gives additional details of this learning rule.

One effect of this weight change is that if the same pattern is presented again and again, over many trials, it will come to evoke a “bubble” of activation centered on X. The fact that nearby nodes are encouraged to respond to similar patterns means that a topographic map eventually develops in which similar inputs are represented by nearby nodes in the network. In this case, the network may be termed a **self-organizing feature map**, since it forms a topographic representation based on some features of the input. Although such a map may be similar to the cortical maps shown in figures 8.3 and 8.5, it is important to note that the network topography depends purely on the frequency and overlap of different input patterns rather than any predefined idea of similarity that may make sense to a human observer.

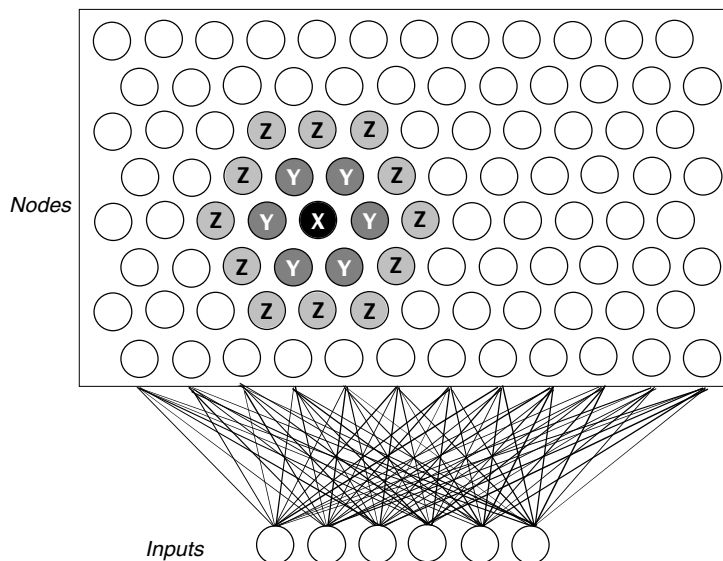


Figure 8.7 A self-organizing feature map is often a network with many nodes arranged in a two-dimensional array. Each node receives weighted connections from all of the input, and these weights are initialized randomly. When an input pattern is presented, one node (X) will become maximally active; it wins the competition. The weights from active inputs to X are strengthened, which means that next time this input pattern is presented, X will respond even more strongly and will be even more likely to win the competition. The weights of X's neighbors are also strengthened, the magnitude of weight change decreasing with distance. Thus, immediate neighbors (Y) will receive a moderate weight increase, more distant neighbors (Z) may receive a mild weight increase, and beyond some neighborhood, there will be no weight changes—or there will even be a mild weight decrease. Over many trials, the input that most strongly activates X will also partially activate immediate neighbors (Y) and mildly activate distant neighbors (Z)—so that a bubble of activation is evoked. Depending on the number of nodes and the size of the neighborhood used, a topographic map will develop in which nearby nodes tend to respond to similar input patterns.

To better understand how these sort of networks function, it is instructive to describe in detail one example from the work of Teuvo Kohonen and his colleagues, who have made extensive study of self-organizing feature maps. They have proposed a “neural phonetic typewriter”¹¹ that learns to transcribe spoken words into output and could be used to generate typewritten text. (For the interested reader, MathBox 8.2 gives some specifics of speech analysis.)

Because Kohonen and many of his collaborators are from Finland, they considered the Finnish language. Finnish has the convenient property of being orthographically regular, which means that words are spelled

MathBox 8.1 Self-Organizing Feature Maps

A self-organizing network learns to represent input patterns on the basis of their structure and frequency, rather than according to any external classification, such as whether they predict reinforcement (e.g., a US). The canonical form of a self-organizing network consists of a single layer of nodes, often arranged as a two-dimensional array (see figure 8.7). Each node has a weighted connection from every element in the input pattern. The weights are initialized to small random numbers in the range 0...1 and may be normalized, meaning that the sum of all the weights coming into each node is the same for all nodes in the network.

When an input pattern is presented, each node j calculates its activation V_j as a weighted sum over the n inputs:

$$V_j = \frac{\sum_i o_i w_{ij}}{n} \quad (8.1)$$

where o_i represents the activation of the i th input and w_{ij} is the weight from that input to j . Since the inputs take on values from 0 to 1, V_j can range from 0 to 1. V_j will be maximized if the n weights coming into j have the same values as the n elements of the input pattern.

Once activation is calculated for all nodes, the winning node w is chosen as the node with maximal activation. (If two or more nodes tie with maximal activation, some probabilistic rule is used to choose between them.) The weights coming into winning node w are updated to increase the correlation between them and the current input pattern, which in turn increases the probability that w will win the competition next time that input pattern is presented. At the same time, nodes that are close to w in the array are updated, and the magnitude of weight change decreases with increasing distance from w . This is done by

calculating the distance from w for each node, d_{wj} (with $d_{ww} = 0$), and then making the magnitude of weight change decrease as d_{wj} increases. A simple learning rule to accomplish this is:

$$\Delta w_{ij} = \beta(o_i - w_{ij})f(d_{wj}) \quad (8.2)$$

where $f(x) \rightarrow 0$ as x gets sufficiently large. β is a learning rate, set to a small positive value. In some cases, β is initially set to a larger value, to allow the net to quickly map out rough organization, and is then slowly decreased to allow progressively finer-grained changes.

Over many training trials, this learning rule has the effect that nearby nodes will tend to respond maximally to similar input patterns, while distant nodes will tend to different kinds of input patterns. This will lead to a rough topography of the kind shown in figure 8.8. Different initial conditions (i.e., different numbers of nodes, different initial weights, and different learning rules) will lead to different topographic maps.

If the number of input patterns is not equal to the number of nodes, then some sharing will occur. In the example of speech recognition (figure 8.8), phonemes that occur most often (e.g., /a/ in Finnish) tend to be represented over several nodes, each of which responds maximally to one variant of the phoneme; infrequent phonemes (e.g., /d/) may be represented by a few nodes, each of which responds to many variants of one phoneme—or even to several similar phonemes.

The learning rules presented here are only examples; many other variations of self-organizing networks have been studied.¹² Self-organizing feature maps have been used in a wide variety of applications in which it is useful to use a computer algorithm to discover underlying order in a large database.¹³

MathBox 8.2 A Brief Introduction to Speech

Speech consists of a continuous stream of acoustic signal. Within any language, this signal can be broken down into meaningful sounds called **phonemes**. For example, the spoken word *cat* is composed of three phonemes: /k/, /æ/, /t/. The phonemic representation of a spoken word often bears no close relationship to the written spelling, especially in a language such as English; for example, the spoken word *knight* would be represented as /nayt/. Different languages recognize different phonemes as significant; for example, /l/ and /r/ are not distinguished in Japanese, while the German /x/ (the guttural final sound in *Bach*) does not occur in English. English uses 38 phonemes; all the languages in the world use a total of about 200.

Phonemes can be defined by their manner of production. Thus, both /t/ and /d/ are generated by forcing through the vocal tract while pushing the tongue against the roof of the mouth just behind the teeth; the difference between /t/ and

/d/ is that the vocal cords vibrate during production of a /d/ but not a /t/. (You can feel this by putting a finger against your throat while uttering each phoneme.) The phonemes /k/ and /g/ stand in a similar relationship, except that the tongue is pressed against the roof of the mouth nearer the back. Phonemes that involve briefly blocking off the passage of air (usually by the tongue or lips) are called consonants; phonemes for which the vocal tract is unobstructed are called vowels.

Each phoneme in a language is associated with a cluster of sound events called **phones**; thus, in English, the first sound in the word “took” and the last sound in the word “out” are both phonemic /t/, even though the actual acoustic sounds (phones) are somewhat different. A native speaker of English will recognize both as the same meaningful unit: “t”.

Each phone generates acoustic energy patterns at different frequencies. This can be plotted as a **speech spectrogram** (see figure 8.A). The vertical

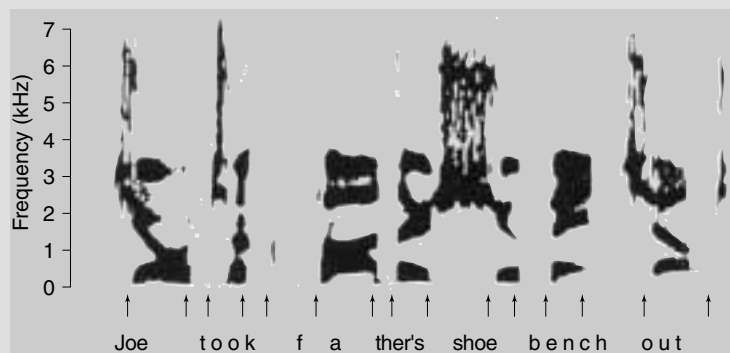


Figure 8.A A speech spectrogram of an English sentence. The vertical axis plots sound frequency, and the degree of darkness indicates amount of acoustic energy at that frequency. The horizontal axis plots time; the sentence being uttered is spelled out below the axis. Most vowels (e.g., /o/, /a/, /e/) consist of several bands of energy that are relatively steady across time, while consonants contain variable onset times and rapid transitions of frequencies. There are also often brief periods of silence (indicated by arrows) during transitions between consonants such as /b/, /p/, /t/, /d/, /k/, /g/, and so these phonemes are often called stop consonants. Although voice, intonation, and speed can vary across and even within speakers, the relative patterns of energy are constant enough that phonemes can be reliably recognized. (Reprinted from Zigmond et al., 1999, p. 1494, figure 57.6.)

MathBox 8.2 (Continued)

axis represents different sound frequencies, and the horizontal axis represents time; the degree of darkness represents the amount of energy present at a given frequency at a given time. Notice that there is no obvious break in energy between words or even between syllables. Instead, there is often a brief period of silence during transitions involving **stop consonants** such as /t/ and /f/, indicated by arrows in the figure. The lack of obvious segmentation between words is one reason why it is so difficult for computers to recognize speech; yet most native speakers of a language perform this task easily.

The precise pattern of acoustic energy generated by a phone will differ with variations in

speaker voice; even within a single speaker, changes in intonation and even nearby phones can affect the spectrogram. Thus, in figure A, the /t/ in *took* generates a somewhat different pattern than the /t/ in *out*, because nearby phones affect pronunciation. Nonetheless, phones—and associated phonemes—can be consistently identified on the basis of the relative combinations of active frequencies of energy. Figure 8.B shows speech spectrograms for the vowels /a/ and /æ/ and for the consonant-vowel syllables /ba/ and /da/.

Vowels such as /a/ and /æ/ tend to produce long-lasting, relatively stable energy at several frequencies. Consonants show rapid transitions. For example, the consonant-vowel combinations

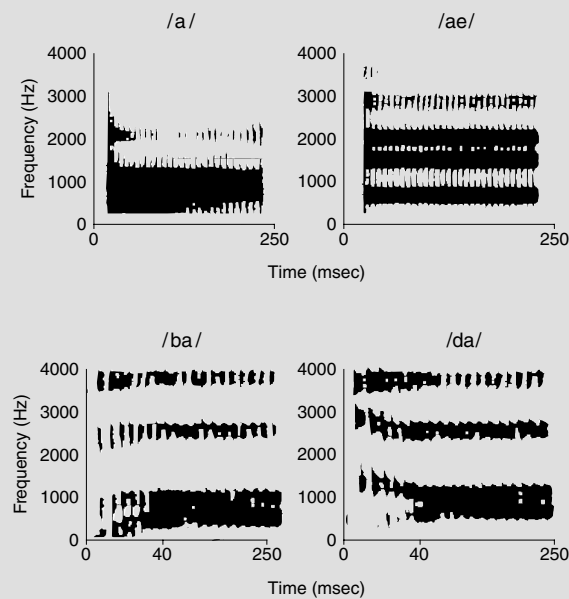


Figure 8.B Spectrographs for the vowels /a/ (as in *father*) and /æ/ (as in *bad*) and for the consonant-vowel syllables /ba/ and /da/. The vowels consist of relatively steady-state activation across a range of frequencies that may last hundreds of milliseconds. Consonants, by contrast, contain variable onset times and rapid transitions of frequencies that may last only a few tens of milliseconds. Distinguishing the /ba/ and /da/ syllables therefore requires discriminating the very rapid transitions that differ between /b/ and /d/. (Reprinted from Fitch, Miller, & Tallal, 1997, figures 1 and 2.)

MathBox 8.2 *(Concluded)*

/da/ and /ba/ differ only in the initial 40-msec segment. This can make them very difficult to discriminate, yet most native speakers of a language make similar discriminations with little conscious effort.

A continuous stream of speech can be converted into a spectrogram, and the spectrogram can then be divided up into short sequences, each of which represents the pronunciation of a single phone, similar to the vowel spectrograms above. Each phone can then be represented as a string of numbers representing the amount of acoustic energy at each frequency. The result is a pattern representing the phone. Changes in speaker inflection and

other variations mean that different instances of the same phone may generate different patterns.

A neural network can be trained to take these input patterns and generate output classifying them as phonemes they represent.¹⁴ The output of such a network could then be translated into typewritten text.

In practice, there may be a considerable degree of additional preprocessing before raw speech signals are translated into input patterns; discussion of such preprocessing strategies is beyond the scope of this book, but the interested reader may refer to specialized publications such as Ramachandran & Mammone, 1995.

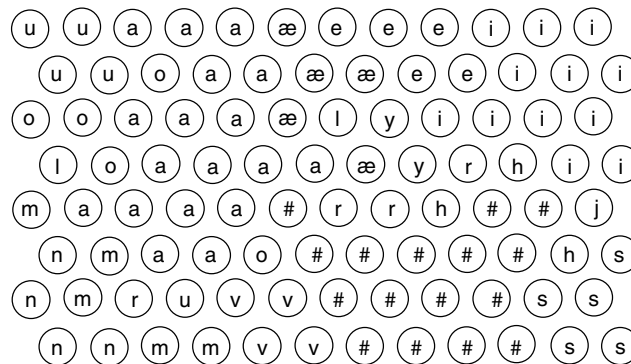


Figure 8.8 The nodes of the phonetic typewriter, labeled with the symbols of the phonemes for which they give best responses. Kohonen and Hari (1999) give the pronunciation guide as follows: /a/ as in *cut*, /æ/ as in *cat*, /e/ as in *bet*, /i/ as in *bit*, /o/ as in *pot*, /u/ as in *put*, and /y/ as in the French word *duc*; /#/ means /k/, /p/, or /t/. (Note that these mappings vary slightly from formal linguistic definitions of phonemes.) To a large extent, similar phonemes are represented by nearby nodes. Therefore, the vowels are grouped together: /m/ and /n/—both involving nasal resonance—are nearby, and so on. Additionally, the amount of space devoted to a phoneme generally reflects frequency; the phoneme /a/, which is the most common in the Finnish language, maps to fully 18 of the 96 nodes. (Adapted from Kohonen & Hari, 1999, figure 2.)

approximately the way they sound. In contrast, English, with its changeable vowels (e.g., the “i” sound differs in “fin” and “fine”) and its strange consonant clusters (e.g., “knight”), has many orthographic irregularities.

Kohonen and colleagues started with recordings of natural speech, which they converted into 15-element input patterns, each representing one speech sound or **phoneme**. These patterns were provided as input to a self-organizing feature map that had 96 nodes arranged in a two-dimensional array. Each node in the network competed to respond to every input, with the winning nodes strengthening their own weights as well as those of their immediate neighbors. After much training, each node in the network came to respond maximally to a preferred phoneme. Figure 8.8 shows the nodes in the network, with each node labeled according to its preferred phoneme.* Several characteristics are evident. First, nodes responding to a particular sound tend to be clustered together. Second, clusters responding to similar sounds tend to

*Note that since this network was trained on examples of Finnish speech, the set of phonemes is not quite the same as a set of phonemes from English speech (Finnish has relatively few fricatives such as sh-, ch-, and th-). Also, note that the node labels shown in the example of figure 8.8 do not map exactly to formal phoneme definitions: The /o/ phoneme is usually taken to be more like the “oh” sound in “hotel” than the “ah” sound in “pot.” The labels are for illustration only.

be close together: The nodes responding to /æ/ (as in “cat”) and /e/ (as in “bet”) are nearby, reflecting the high similarity in sounds; the nodes responding to /m/ and /n/—two consonants that involve nasal resonance—are neighbors; and so on. More generally, vowels are gathered near the top half of the map, while consonants are gathered in the bottom half. The topography is not perfect; for example, the nodes responding to /l/ are quite widely distributed across the network, while the highly similar consonants /p/, /t/, and /k/ are not well distinguished (represented as /#/ in figure 8.8). Nevertheless, considering that the network was given no supervision on how to encode phonemes, the map does quite a good job of representing the acoustic similarities between sounds. Finally, there is rough correspondence between the importance of phonemes and the amount of total area representing them: /a/, which is the most common Finnish phoneme, activates fully 18 of the 96 nodes, while the less common /j/ activates only one node maximally.

These emergent properties of the self-organizing map are quite reminiscent of the topographic maps that are encountered in primary sensory cortex (refer to figures 8.3 and 8.5): In each case, cells evolve that preferentially respond to particular types of input, and neighboring cells tend to respond to similar inputs. Thus, this class of network is a possible model for how topographic maps could develop in the brain during early exposure to a set of stimuli. Of course, there may be many ways such maps could emerge, and self-organizing networks instantiate only one of these possibilities. However, many of the assumptions inherent in such a network are consistent with features of cortical anatomy and physiology, a topic to which we turn next.

Piriform Cortex

Most parts of cerebral cortex in mammals are six-layered neocortex (figure 8.1A). However, there exist a few places in the brain where the cortex begins to change, gradually becoming two-layer paleocortex similar to the reptile cortex shown in figure 8.1B. One such area is the olfactory cortex (also known as **piriform cortex**). The hippocampus itself is also a two-layered sheet of paleocortex, although it is folded in on itself to create its characteristic “C” shape. The piriform cortex abuts the hippocampal region; in some species such as the rat, the piriform cortex actually merges into nearby cortical areas, and it is therefore difficult to make a precise division between the two regions. One theory is that hippocampus and piriform cortex are remnants of the “primitive brain” from which more complex neocortex evolved.

Whatever its history, the relatively simple organization of the piriform cortex suggests that it is a logical place to begin exploring how cortical

anatomy could give rise to computational function. Additionally, although human olfactory cortex is relatively small, the olfactory cortex of most mammals is much larger, reflecting the importance of odor information to those species. Dogs, for example, have almost fifty times as many olfactory neurons as humans do. Therefore, the olfactory cortex is especially convenient for animal studies.

Olfaction (the sense of smell) begins with a thin sheet of olfactory receptor cells, high up in the nasal cavity shown in figure 8.9, called the olfactory epithelium. It contains olfactory receptor cells that transduce chemical stimuli (odors) into nerve impulses by processes that are still incompletely understood.

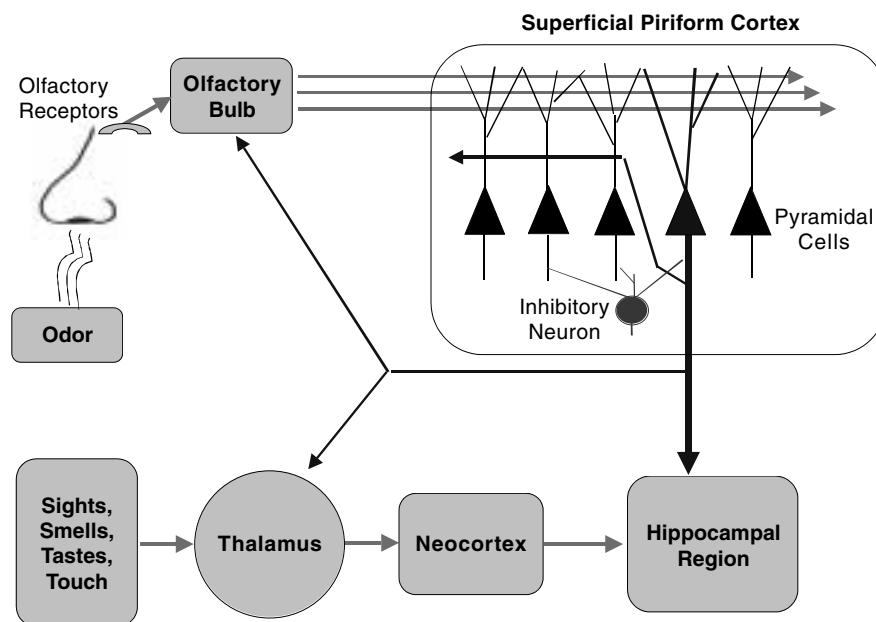


Figure 8.9 Schematic of olfaction. Odors enter the nose and reach olfactory receptors in the olfactory epithelium, which in turn transmit information to the olfactory bulb and thence to the olfactory cortex. Piriform cortex is one subarea of olfactory cortex. Pyramidal cells in the piriform cortex in turn project to other pyramidal cells as well as to local inhibitory neurons, which inhibit activity in nearby pyramidal cells. Pyramidal cells also project out of piriform cortex to thalamus, hippocampal region, and back to olfactory bulb. Olfaction is unique in that it is the only sensory modality in which sensory inputs project directly to cortex; most other sensory information travels through thalamus before reaching neocortex. Piriform cortex is also the only primary sensory cortical area with substantial direct projections to the hippocampal region.

The axons of the olfactory receptor cells form the olfactory nerve and travel to the **olfactory bulb**; from there, information projects to the olfactory or piriform cortex. This arrangement is special: No other sensory information has direct access to cortex. Most other sensory information travels first to thalamus and then to primary sensory areas in neocortex. The direct access of odors to cortex means that it should be possible to model the function of piriform cortex in some detail without first requiring a full understanding of thalamic preprocessing.

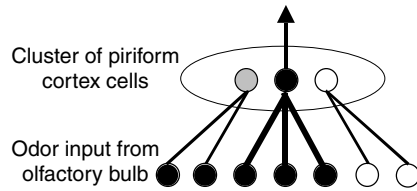
A final special characteristic of piriform cortex is that it is the only sensory cortex with a sizable direct connection to the hippocampal region; all other sensory information passes through several layers of unimodal and polymodal association cortex before reaching the hippocampal region.¹⁵

Several researchers have presented computational models of olfaction;¹⁶ Richard Granger, Gary Lynch, and colleagues proposed a competitive learning system that includes many of the known anatomical and physiological properties of piriform cortex.¹⁷ Their model includes a learning rule that incorporates what is known about plasticity in piriform cortex and model neurons that incorporate known detail about piriform neuronal activity. However, many of the important properties of the system can be understood by using the simple example shown in figure 8.10.

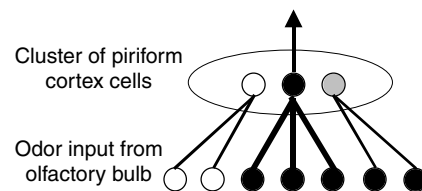
The network model in figure 8.10 consists of two interacting systems: a layer of nodes representing the olfactory bulb and a layer representing pyramidal cells in superficial piriform cortex. These pyramidal nodes are organized into small clusters (typically 10–20 nodes per cluster), each of which functions as a small competitive network. When an olfactory input arrives, the nodes in a cluster are differentially activated. One node in each cluster is most strongly activated; this node wins the competition and outputs its response, while feedback inhibition silences the nonwinning nodes (figure 8.10A). One of the unique features of the piriform model is that the input classification does not stop here. The piriform activity feeds back to inhibit the olfactory bulb, specifically suppressing the parts of the input pattern that activated the winning nodes (figure 8.10B). This allows new parts of the input pattern to activate new piriform nodes, which win the second round of the competition. The entire process may be repeated a number of times until all the olfactory input is suppressed and the network becomes quiescent.

Now suppose another input is presented that shares some features with the first odor (figure 8.10C); the common features will tend to activate the same initial response. Once these are masked out, the secondary response may be different (figure 8.10D). Thus, this network is capable of performing **hierarchical clustering** of inputs. The initial response to a stimulus is a coarse

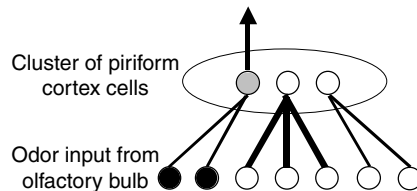
(A) Input Pattern A: Step 1



(C) Input Pattern B: Step 1



(B) Input Pattern A: Step 2



(D) Input Pattern B: Step 2

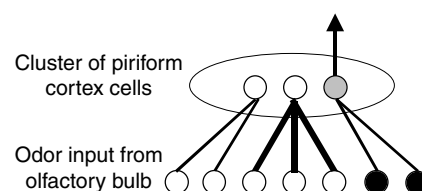


Figure 8.10 Simplified version of the operation of the piriform clustering model (Ambros-Ingerson, Granger, & Lynch, 1990). A series of inputs, representing odor information from olfactory bulb, connect to a large number of piriform cortex cells. Piriform cortex cells are organized into small clusters (one cluster is shown). (A) A familiar odor is presented to the olfactory bulb and elicits a pattern of activation in each piriform cortex cluster. Piriform nodes compete, the most strongly active node winning: It outputs maximally, and the other nodes are silenced. (B) Piriform activity feeds back to the bulb, masking out the portions of the input that activated the winning nodes. This allows the remaining input features to activate new piriform nodes on the next iteration. Many iterations may occur, each producing a different spatial activity pattern in the clusters. (C) A new odor input is presented that shares some features with the input in A. One piriform node wins the competition—the same as in A. This represents a coarse clustering, classifying the two odors into the same broad category of input. (D) After feedback masks the odor input, however, the remaining features activate a new piriform node. Thus, a finer classification is made that differentiates the current odor from the odor in A.

classification, which may not distinguish between several similar inputs. However, successive responses do become more and more fine-grained.

Conceptually, the network might first recognize a food odor, then recognize that this particular input is a cheese odor, and eventually identify a specific type of cheese. What is most interesting about this network is that this sophisticated analysis of an input in terms of successively finer-grained classification occurs without any external teaching signal; the network simply uses competitive learning to group similar inputs together.

There is a second important principle at work in the piriform cortex model: representational compression. If two stimuli co-occur, the network has no way of knowing this *a priori*. Instead, it will treat the two stimuli as a single, compound stimulus (figure 8.11A). For example, co-occurring floral and musky odors would be treated as a single odor with components of both floral and musky scents. The network will form a cluster in response to these features. If one of the component odors (e.g., floral scent) later appears alone, the network will recognize this as a degraded version of the compound odor, the network will recognize this as a degraded version of the compound odor,

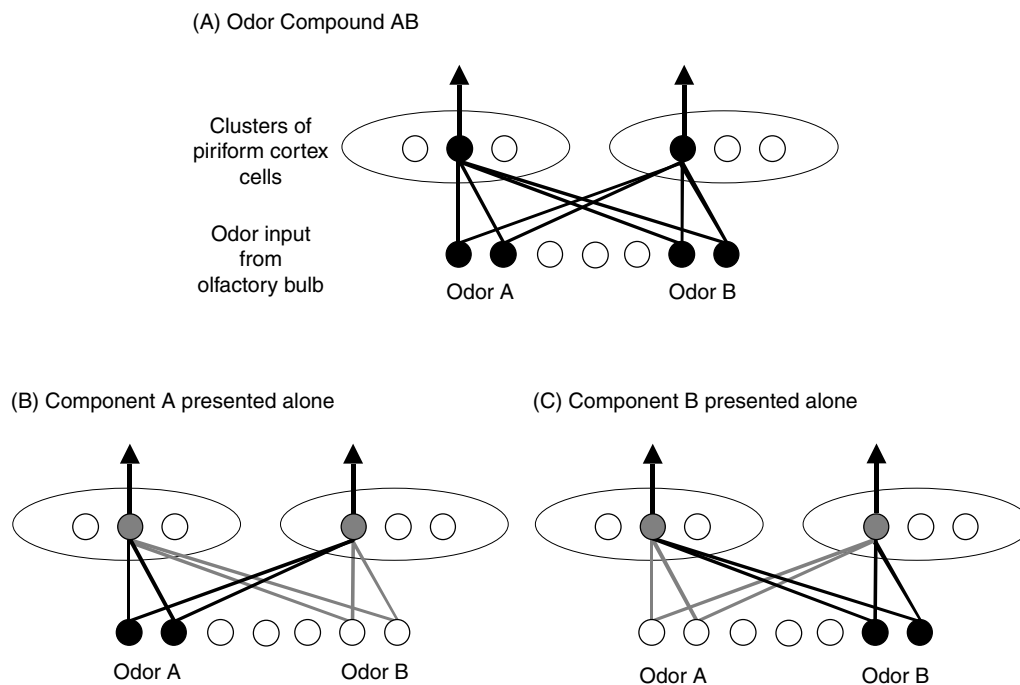


Figure 8.11 Co-occurring odors are compressed in the piriform model. (A) The network is trained on an input pattern that consists of two odor components A (e.g., a floral scent) and B (e.g., a musky scent). Within each node cluster (two are shown), one node wins the competition to respond to this compound odor, and the weights from active inputs to those nodes are strengthened. For simplicity, only strong connections are shown. (B) Later, one of the component odors is presented alone. This input will tend to activate the same nodes as the compound, along previously strengthened connections. The nodes will be less strongly activated by the individual odor, but as long as their activation is greater than that of any other node in the cluster, they will still win the competition. Thus, the network will tend to respond to the component odor in the same way as it did to the compound. (C) If component B is presented alone, the same response occurs. In effect, the representations of co-occurring odors A and B have been compressed, so the response to either component is similar.

and it will respond by activating the nodes coding that compound odor (figure 8.11B). *Thus, Granger and Lynch's model compresses the representations of co-occurring odors.*

From a computational perspective, this representational compression is essentially the same as that which occurs in our own cortico-hippocampal model.¹⁸ The difference is that, while the hippocampal region processes a full range of polymodal information, the piriform cortex is limited (mainly) to processing odor stimuli. Furthermore, whereas the hippocampal region is hypothesized to both compress and differentiate representations as appropriate, the piriform model is capable only of representational compression.

This difference has several important implications, the most important of which is that once co-occurring odors are compressed, there is no way for the piriform model—working in isolation—to undo this compression. Thus, if odors A and B originally occur together (as in figure 8.11) and later need to be mapped to different responses, this mapping will be very difficult.

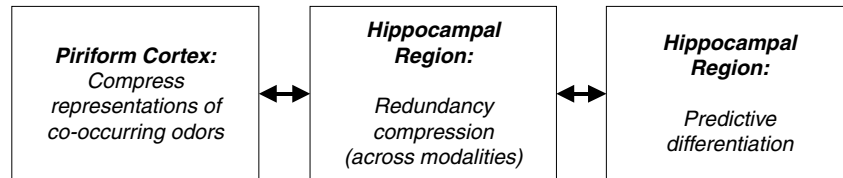
Of course, normal animals can learn to distinguish odors (and other stimuli) that have previously co-occurred. Apparently, some other mechanism is available to redifferentiate representations as needed. One possibility is the hippocampal region.

Piriform-Hippocampal Interactions

Our cortico-hippocampal model assumes that the hippocampal region can form new stimulus representations that compress redundant information while differentiating predictive information. Normally, the hippocampus receives highly processed, multimodal information from a whole range of stimulus modalities, and thus its representations can take into account the cross-modal features of individual stimuli or the co-occurrence of stimuli across different modalities. Sensory cortex, particularly primary sensory cortex, is largely limited to processing information in a single sensory modality. Thus, the piriform network described in the preceding section may compress the representations of co-occurring odors but not stimuli in other modalities. It is also unable to redifferentiate representations once they have been compressed.

One way to schematize the relationship between piriform cortex and hippocampal region is shown in figure 8.12A. Odor inputs are first processed by piriform cortex, which forms odor clusters and also compresses the representations of co-occurring odors. Because the piriform cortex has direct connections to the hippocampal region, these new odor representations are combined with highly processed information from other modalities. This, in

(A) Possible Normal Piriform–Hippocampal Region Interaction



(B) Effects of Hippocampal-Region Damage

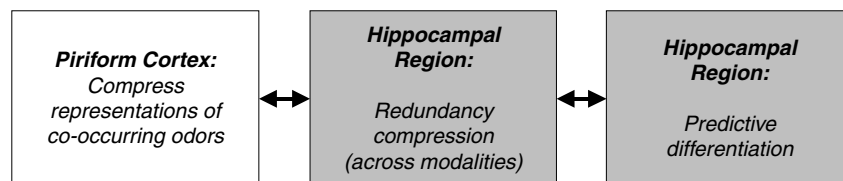


Figure 8.12 Possible interactions between piriform cortex and hippocampal region. (A) Normally, the piriform cortex may tend to compress the representations of co-occurring odors. The hippocampal region also performs compression of stimuli that co-occur or are otherwise redundant; this may especially involve compression across stimulus modalities. The hippocampal region also performs predictive differentiation, pulling apart the representations of stimuli that are differentially predictive of future reinforcement. This serves as an opponent function for representational compression. Normally, the two processes are in balance. (B) Hippocampal-region damage removes the ability to perform cross-modal compression and the ability to perform predictive differentiation. The result is that undamaged compression processes in piriform cortex are allowed to proceed unchecked; overall, this will lead to a tendency to overcompress stimulus information, making it hard to distinguish stimuli which co-occur.

turn, generates new representations that compress redundant information while differentiating predictive information. In this case, hippocampal-region damage does not absolutely eliminate the ability to alter stimulus representations (figure 8.12B). The piriform cortex can still alter the representations of odor stimuli, as, presumably, other cortical areas do within their respective domains. But two abilities are lost. First, without the hippocampal region to integrate across modalities, there is relatively little ability to combine information across different sensory domains. Second, there is relatively little ability to differentiate representations, particularly representations that have previously been compressed. The net result is that cortical compression proceeds without the opponent process of differentiation, and, at least within a single sensory domain, there is a tendency to overcompress information,

which in turn hinders subsequent differentiation. For the piriform cortex, this means that it will be very difficult to distinguish individual scents that have previously been presented together. Thus, we might expect that without hippocampal-dependent differentiation (due to brain damage), an animal might tend to overcompress or fuse stimuli that have appeared together in a compound.

In fact, exactly this kind of impairment *has* been observed in odor discrimination studies done in rats with hippocampal-region damage. Howard Eichenbaum and colleagues have extensively studied rat odor discrimination. In their laboratory, a rat is placed in a small chamber with a recessed port from which an odor is delivered as shown in figure 8.13A. Typically, one odor (e.g., A) is defined as positive. The rat's task is to learn to respond to A by pushing its nose into the port delivering A and holding this position for a few seconds. This behavior is rewarded by a small food or water reward. Then an elaborate vacuum system removes evidence of both odors from the chamber, and new odors are delivered for a new trial.

When a negative odor (e.g., B) is delivered, the rat must sample the odor, then withdraw its nose from the odor port to initiate the next trial. This kind of task is called **successive odor discrimination**, and normal rats can learn fairly easily, even if a large number of odors are used. Rats with a fornix

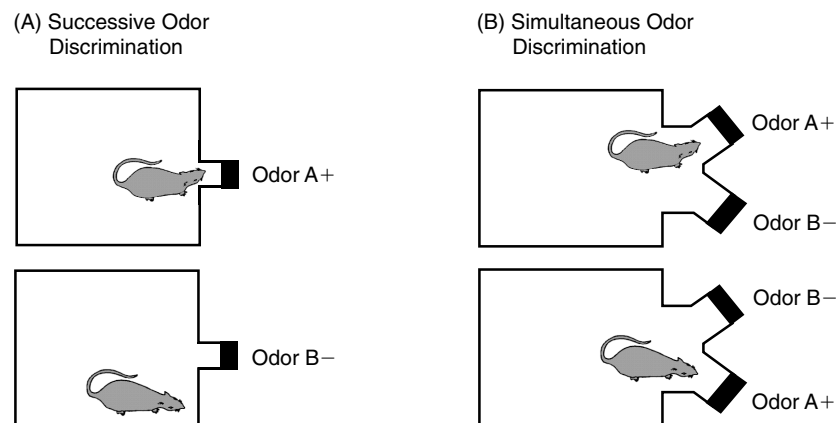


Figure 8.13 (A) Successive odor discrimination paradigm. The rat is presented with one of two odors from an odor port and must learn to poke its nose into the odor port in response to the rewarded odor A but not the nonrewarded odor B. (B) Simultaneous odor discrimination paradigm. Two odors arrive simultaneously; the rat must learn to nosepoke to the rewarded odor, regardless of spatial position. Hippocampal-region damage spares successive odor discrimination but impairs simultaneous odor discrimination (Eichenbaum et al., 1988). (Adapted from Myers & Gluck, 1996, figure 2.)

lesion (which severs an important pathway into and out of hippocampus) are also generally unimpaired at learning successive odor discriminations.¹⁹

Howard Eichenbaum and colleagues have considered a variation of this paradigm in which there are two odor ports that can simultaneously deliver separate odors (figure 8.13B); the rat must then nosepoke to the left-hand or right-hand odor port, depending on which port is delivering the positive odor. For example, given odor A+ at the left port and odor B− at the right port, the rat should nosepoke left to obtain water; given odor A+ at the right port and odor B− at the left port, the rat should nosepoke right to obtain water. This kind of task, called **simultaneous odor discrimination**, is severely impaired by fornix lesion.²⁰ Entorhinal lesion also disrupts retention of a similar task.²¹ The lesioned animals' impairment cannot be ascribed to failure to nosepoke left or right, since under other circumstances, they can learn to do this;²² therefore, the critical difference between the simultaneous and successive discrimination paradigms seems to be whether odors are presented singly or in pairs. The lesioned rats can cope with singly presented odors but not with odor pairs.

The schematized model shown in figure 8.12 suggests why this might be so. When odors are presented singly, as in successive discrimination, the piriform cortex will develop a representation for each; when odors are presented in compound, the piriform network will tend to compress representations of those component odors, making it harder to distinguish them—and therefore harder to learn to respond to one but not the other. Eichenbaum and colleagues have similarly suggested that hippocampal-region damage leads to a tendency to fuse representations of co-occurring odors.²³

There is some evidence of this kind of overcompression or fusion deficit following hippocampal-region damage in other species. For example, retention and relearning of simultaneously presented odors is disrupted in monkeys with hippocampal-region damage.²⁴ It is less clear whether stimuli from other modalities are disrupted the same way. Thus, after hippocampal-region damage, choosing between two simultaneously presented visual cues is sometimes disrupted in monkeys and sometimes spared.²⁵ The data are similarly mixed for tactile and auditory discrimination in various species.

One possible explanation for these mixed results is that odor stimuli are, by their very nature, particularly prone to mix and blend, and therefore, co-occurring odors are especially subject to compression.²⁶ Visual stimuli, in contrast, may be prone to compression only if they are presented close together or if they are different features of the same object, such as color and form. Under these conditions, visual stimuli might be more likely to show compression—and more likely to show overcompression after hippocampal-region damage. This idea remains to be explored experimentally.

Integrating Piriform Cortex with the Cortico-Hippocampal Model

The idea that the piriform cortex compresses co-occurring odor stimuli while the hippocampal region compresses and differentiates all kinds of stimuli is relatively straightforward to instantiate within the general framework of our cortico-hippocampal model.²⁷ Two basic changes are required. The first, obviously, is the inclusion of a piriform network to perform preprocessing of odor stimuli. Nonodor inputs do not pass through this network; they are presumably preprocessed in other, nonolfactory cortical areas that are not included in the model (figure 8.14A). The outputs of the piriform network pass directly to the hippocampal-region network, consistent with the anatomical evidence.

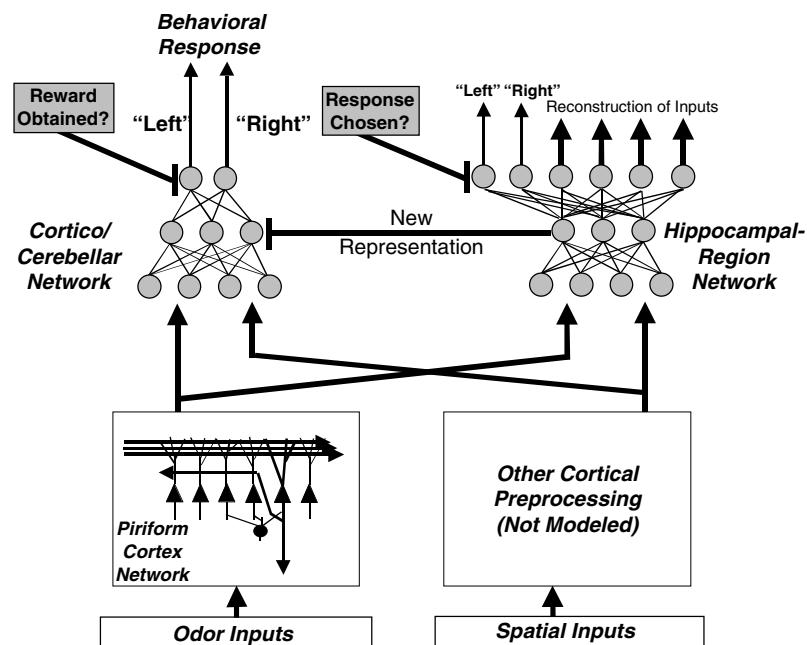


Figure 8.14 (A) The cortico-hippocampal model extended to allow odor preprocessing with a piriform network. Inputs encode odor quality and spatial arrangement. Odor inputs are preprocessed by the piriform cortex network, which clusters odors and compresses co-occurring odors. Other cortical areas may similarly preprocess spatial information, but this is not simulated here. The preprocessed inputs are provided to the cortico-hippocampal model. The cortico/cerebellar network learns to map from these inputs to a behavioral response (choose left or right). The hippocampal-region network learns to reconstruct its inputs and predict the behavioral response. As usual, new stimulus representations constructed in the hippocampal-region network are adopted by the cortico/cerebellar network. Full details of this model are given in the appendix to this chapter.

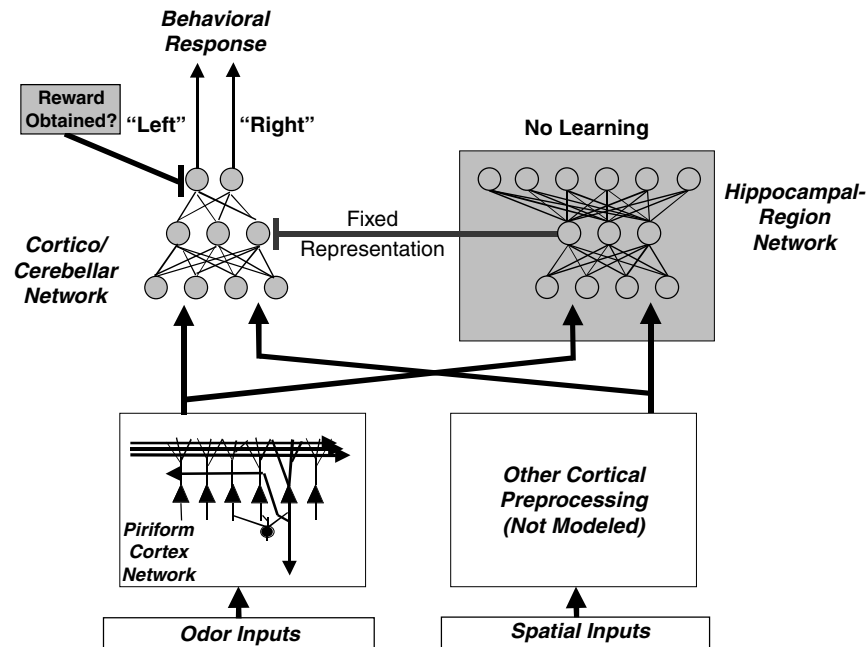


Figure 8.14 (B) The fornix-lesion model. Fornix lesion does not actually involve direct damage to the hippocampus but severs an important modulatory pathway connecting the hippocampus with structures such as thalamus and medial septum. Evidence suggests that this pathway may enable learning in the hippocampus. Thus, fornix lesion is modeled by eliminating learning in the hippocampal-region network. Information still passes through the network and is transformed into a representation in the internal node layer, according to random, fixed weighted connections; this representation is not subject to the usual hippocampal-mediated constraints of redundancy compression and predictive differentiation and instead is a random recoding of the input information. The cortico/cerebellar network must learn to map from these fixed representations to appropriate behavioral output. Full details of this model are given in the appendix to this chapter.

Second, unlike the prior application of our model to classical conditioning, in which the response was a single behavioral action (e.g., blink), in simultaneous odor discrimination the response is a choice between two or more responses (e.g., nosepoke left versus nosepoke right).

The cortico/cerebellar network (see figure 6.4 or figure 7.9) is easily generalized by giving it multiple output nodes, one representing each possible response. The hippocampal-region network is adapted to produce output that reconstructs its inputs and also predicts the behavioral response selected. This system can now be applied to simultaneous discrimination tasks such as those studied by Eichenbaum and colleagues.

A hippocampal-region (HR)-lesioned model could be generated by simply disabling the hippocampal-region network. In this case, odor inputs would be preprocessed by the piriform network and then project directly to the cortico/cerebellar network, as schematized in figure 8.14B.

However, in the relevant behavioral studies, Eichenbaum and colleagues generally did not consider HR lesions but rather fornix lesions. As we described earlier, a fornix lesion does not directly damage the hippocampus. Rather, it interrupts an important pathway connecting the hippocampus with subcortical structures such as the thalamus and septum.²⁸ Under some circumstances, fornix lesions may impair learning as much as outright hippocampal lesion,²⁹ although different lesions can and do have different effects;³⁰ we will return to this issue later, in chapters 9 and 10. For now, we merely note that fornix lesion destroys critical input to the hippocampus that is known to be important for hippocampal learning.³¹

Therefore, to implement a fornix-lesion model, we simply assume that the hippocampal-region network is not physically damaged, but rather is incapable of learning new information. Thus, information passes through the hippocampal-region network as usual, but the weights in the hippocampal-region network never change in response to input (figure 8.14B). The result is that the only representational changes that occur in the fornix-lesion model are those mediated by the piriform network, which tend to overcompress co-occurring odor stimuli. Full implementation details of the intact and fornix-lesion models are given in the appendix to this chapter.

Now the intact and fornix-lesion models can be applied to the simultaneous odor discrimination task shown in figure 8.13B. On each trial, two odors (e.g., A and B) are presented, together with information encoding their spatial location (i.e., left or right port). The piriform network clusters the odor information, and then the hippocampal-region network forms new stimulus representations that include both odor and spatial information. These new representations are adopted by the cortico/cerebellar network, which learns to map from them to the correct behavioral response (i.e., choose left or right). Figure 8.15B shows that the intact model can learn such a simultaneous discrimination quickly. In the intact model, the piriform network clusters the representations of odors A and B, which always co-occur. This information, together with the spatial information, is passed on to the hippocampal-region network. The hippocampal-region network forms new internal representations that differentiate the information needed to solve the task. In this case, the relevant information is which odors occur in which locations—in other words, whether odor A is on the left or the right. All other information, including the odor cluster

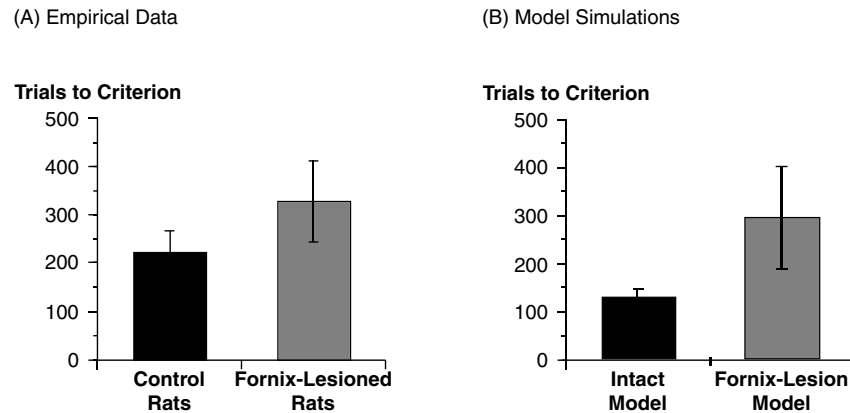


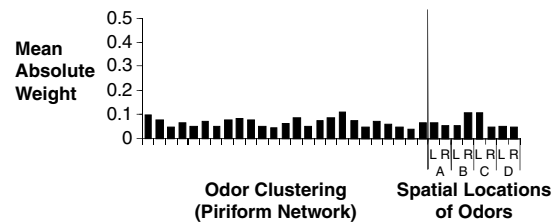
Figure 8.15 Simultaneous odor discrimination: Choose odor A over odor B, regardless of spatial ordering. (A) Control rats learn this task relatively quickly, while fornix-lesioned rats are slower. (Plotted from data presented in Eichenbaum et al., 1988.) (B) The intact and fornix-lesion models show a similar pattern. (Plotted from data presented in Myers & Gluck, 1996.)

information provided by the piriform network, is irrelevant for this end. The hippocampal-region representations therefore come to emphasize spatial information at the expense of the piriform clustering information. Figure 8.16 shows an example of such a representation. Initially, there are small weights to the internal-layer nodes from all inputs, both those from the piriform network and those detailing spatial information (figure 8.16A). By the end of learning, however, the weights containing information about the spatial locations of A and B (that is, whether each is arriving from the left or right port) have become heavily weighted (figure 8.16B); the weights from the piriform network—as well as those devoted to other odors such as C and D—remain low.

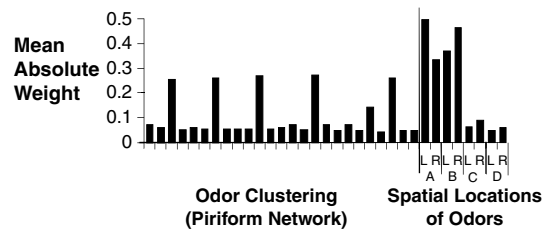
Figure 8.16 Magnitude of weights from each input to the internal-layer nodes in the cortico/cerebellar network, averaged across all internal-layer nodes. (A) In the intact model, all weights are initially random and small. (B) After training on A+, B−, the hippocampal-region network develops new internal representations that differentiate predictive information—namely, odor position—that are acquired by the cortico/cerebellar network. The new representations tend to strongly weight information detailing whether A and B appear on the left or right. To a lesser degree, some odor clustering information from the piriform network may be weighted as well. (C) The fornix-lesion model starts out with initial weights in the cortico/cerebellar model that are much like the initial state of the intact model. (D) However, the representations in the hippocampal-region network are fixed and do not come to emphasize relevant information. These are nonetheless adopted by the cortico/cerebellar network. Since no particular weight is given to predictive inputs, it is very difficult for the cortico/cerebellar network to learn the correct behavioral response on the basis of these representations. (Reprinted from Myers & Gluck, 1996, figure 9A,B,D,E.)

The situation is very different in the fornix-lesion model. The initial weights start out much the same as in the intact model (figure 8.16C), but since there is no hippocampal-region learning, representations are not adapted to emphasize predictive information. The random, fixed representations of the

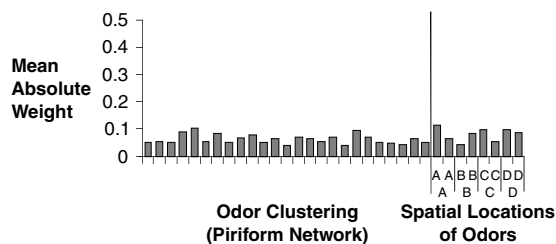
(A) Intact Model: Initial Weights



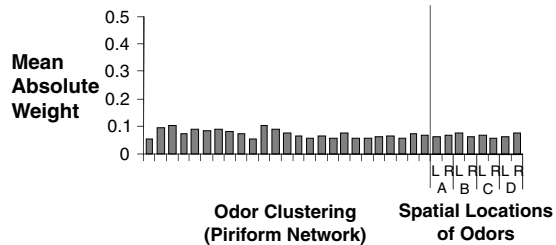
(B) Intact Model: After A+, B- Training



(C) Fornix-Lesion Model: Initial Weights



(D) Fornix-Lesion Model: After A+, B- Training



hippocampal-region network are adopted by the cortico/cerebellar network, but these tend not to contain much useful information (figure 8.16D). This makes it very difficult for the cortico/cerebellar network to map different odor locations to different behavioral responses, resulting in poor overall performance of the fornix-lesion model (figure 8.15B). This is consistent with the generally poor performance of fornix-lesioned rats on simultaneous odor discrimination.³²

The average performance data shown in figure 8.15 hide some important individual differences. Figure 8.17A shows that most intact rats solve a simultaneous discrimination quickly, within a few hundred trials. Figure 8.17B shows similar performance in the intact model. On average, both the fornix-lesioned rats and fornix-lesion model do much worse. However,

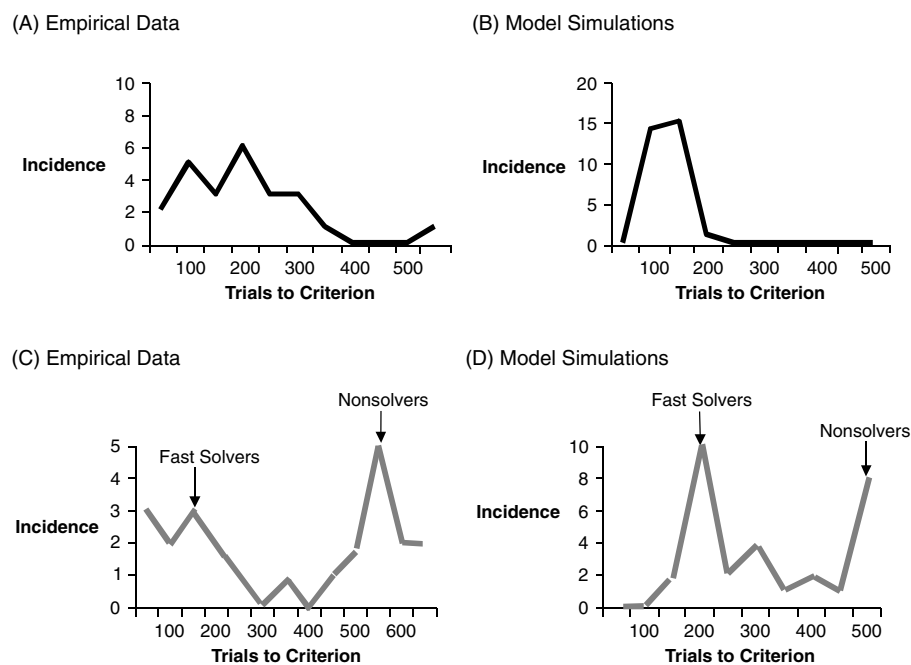


Figure 8.17 Distribution of learning times on simultaneous odor discrimination, in terms of trials to reach criterion, pooled over all tasks. (A) Control rats show a unimodal distribution, solving most tasks within fewer than 300 trials. (B) The intact piriform-hippocampal model shows a similar distribution of solution times. (C) Fornix-lesioned rats show a bimodal distribution, either failing to learn the discrimination within 500 trials or else learning just as quickly as control rats. (D) The lesioned model shows a similar bimodal distribution, most simulations solving tasks in about 200 or more than 500 trials. (Reprinted from Myers & Gluck, 1996; A and C are plotted from data presented in Eichenbaum et al., 1988.)

occasionally and seemingly at random, the fornix-lesioned rats do solve a simultaneous discrimination—and when they do, they do so just as quickly as control rats (figure 8.17C).

The fornix-lesion model shows similar performance, and our model provides an interpretation of the curious bimodality seen in the data from lesioned animals.³³ In our intact model, the hippocampal-region network constructs new stimulus representations that emphasize predictive information. In this case, the result is that internal-layer nodes come to respond strongly to information about the spatial layout of odors and largely ignore the overcompressed information from the piriform network.

Figure 8.18A shows an example of the kind of internal-layer node that tends to develop in the hippocampal-region network and then be adopted by the cortico/cerebellar network. The node is strongly active whenever a positive odor (A or C) is present at the left-hand port and a negative odor (B or D) is present at the right-hand port and a negative odor

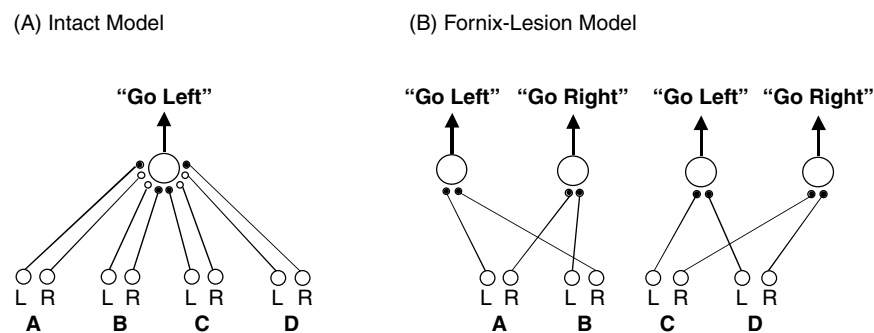


Figure 8.18 Even when an intact simulation and a fornix-lesioned simulation learn the same two problems (A+,B− and C+,D−), they do not learn them in the same way. (A) In the internal layer of the intact model, new representations are formed by the hippocampal-region network. The result is that internal-layer nodes develop such as the one shown here that learn general response rules: Go left whenever a positive stimulus is on the left (A or C) and a negative stimulus is on the right (B or D), and go right otherwise. Many of these nodes will develop in the intact model's internal layer, allowing the problem to be solved quickly. Further, when familiar odors are presented in novel recombinations (e.g., A versus D), this type of node continues to give the correct response. (B) Internal-layer nodes are fixed in the hippocampal-region network of the fornix-lesion model, meaning that nodes like the one in A cannot be constructed. In fact, the only way the fornix-lesion model can solve the discriminations is if initial, random weights fortuitously preserve some information about the spatial location of odors. If this happens at all, it is likely to be only about individual odors. Thus, a fornix-lesioned network that solves the two discriminations might have four internal-layer nodes, each of which responds to one spatial arrangement of one odor pair: go left for AB, go right for BA, go left for CD, go right for DC. This solution works well enough for the initial discriminations but will be of little help when the odors are recombined (e.g., A versus D).

(B or D) is present at the right-hand port; it is strongly inhibited by the reverse placements. Thus, this node should strongly activate the output node associated with a “go-left” response and strongly inhibit the output node associated with a “go-right” response.

The fornix-lesion network cannot construct new stimulus representations in the hippocampal-region network, and it is forced to learn on the basis of preexisting representations that will tend to include a large amount of compressed information from the piriform network. Generally, these preexisting representations will not preserve enough spatial information to allow an arbitrary odor discrimination to be solved. On occasion, however, the random initial weights in the fornix-lesion model’s hippocampal-region network are such that they happen to preserve enough spatial information to solve a particular task. All it may take is a single internal-layer node that responds strongly when A is present on the left and B is on the right and responds weakly for the reverse ordering. Figure 8.18B shows an example of some internal-layer nodes that might exist in a fornix-lesion simulation. One node happens to respond when A is on the left and B is on the right (AB); another happens to respond to the reverse ordering (BA). By strongly weighting connections between these nodes and the output nodes in the cortico/cerebellar network, the model can learn to generate the correct response. If more nodes exist that similarly code CD and DC, the simulation will be able to learn that discrimination as well. In such a fortuitous case, the fornix-lesion model will be able to solve both discriminations—and solve them fairly quickly, as is shown in figure 8.15B.

This interpretation within our computational model is similar to the qualitative explanation suggested by Eichenbaum and colleagues. They argued that the fornix-lesioned rats could solve a discrimination only “as a fortuitous consequence of idiosyncratic perceptual variations.”³⁴

Although the fornix-lesioned rats and fornix-lesion model can occasionally solve a discrimination as quickly as their nonlesioned counterparts, this does not necessarily mean that they are using the same strategies. Eichenbaum and colleagues demonstrated such a difference in rats: Each fornix-lesioned rat was trained on various successive discriminations until the experimenters found two odor pairs (e.g., A+ versus B– and C+ versus D–) the rat could learn. A control rat was then trained on the same two odor pairs. Then the animals were presented with the familiar odors in novel recombinations (e.g., A versus D). Control rats continued to respond to the previously positive odor A instead of the previously negative odor D. Fornix-lesioned rats performed at chance, showing no preference for the previously positive odor, even though they continued to respond correctly when presented with the A versus B discrimination (figure 8.19A).³⁵

(A) Empirical Data

(B) Model Simulations

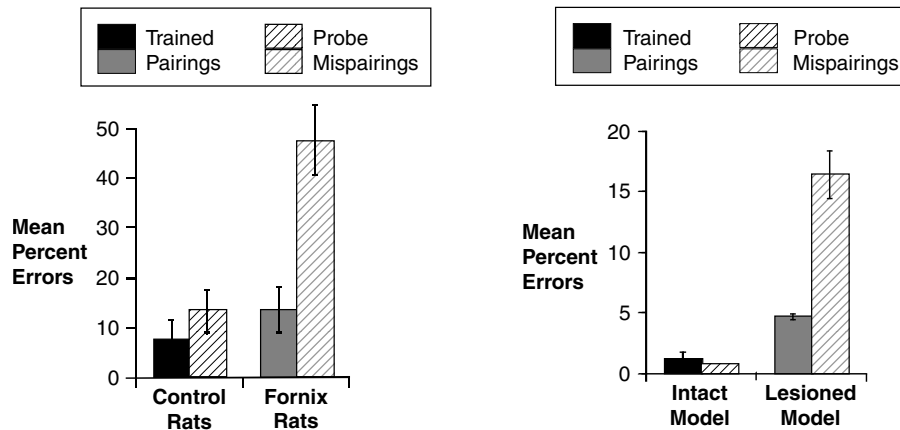


Figure 8.19 (A) After training on the odor discriminations A+ versus B− and C+ versus D−, rats were given test trials consisting of mispairings of the familiar odors: A+ versus D− and C+ versus B−. Control rats continued to perform very well on these probe mispairings. Fornix-lesioned rats, by contrast, performed near chance on these probe mispairings. (Plotted from data presented in Eichenbaum, Mathews, & Cohen, 1989.) (B) The piriform-hippocampal model shows a similar trend: The intact model performs well on both trained pairings and probe mispairings; the lesioned model shows a strong impairment on the probe mispairings. (Adapted from Myers & Gluck, 1996, figure 14.)

The intact and fornix-lesion models show the same effect (figure 8.19B): After training on two pairs that the lesioned model happens to master, the intact model performs nearly perfectly on the transfer test, while the lesioned model does not.³⁶ The internal-layer representations in figure 8.18 show why. When the intact network is presented with a novel recombination of trained stimuli (e.g., A versus D), internal-layer nodes such as the one in figure 8.18A can cope with the recombination: The node will still fire strongly to A on the right and D on the left and weakly to the other spatial arrangement, and therefore it is still useful as a “go-right” node. By contrast, the internal-layer node in figure 8.18B is activated by combinations of A and B and is only very weakly activated by either stimulus alone. Its output will not be very useful in gauging a response to the new combination. Further, because there is no new representational change in the fornix-lesion model’s hippocampal-region network, the situation does not improve even after extensive training on the new combinations.

Thus, our model accounts for the subtle pattern of data seen in Eichenbaum’s studies of olfactory discrimination in both the intact and lesioned rats.³⁷

At the beginning of this section, we noted that the piriform clustering model of Granger and Lynch suggests that sensory cortices can adapt their representations to reflect superficial similarity between inputs. However, what these brain regions may not be able to do is to operate across sensory domains: for example, encoding both the visual and olfactory features of a stimulus into a unified compound percept. The sensory cortices may also not be able to differentiate stimulus representations based on meaning: For example, if both A and B (which are superficially similar cues) differentially predict the US, then their representations should be differentiated to facilitate mapping them to different responses. These kinds of multimodal and meaning-driven representational changes may lie uniquely within the domain of the hippocampal region. The next chapter reviews a model suggesting that a specific hippocampal-region structure, the entorhinal cortex, may be the locus of multimodal redundancy compression.

8.3 RELATIONSHIP OF COMPUTATIONAL MODELS TO QUALITATIVE THEORIES

So far, this chapter has discussed how primary sensory cortex may operate as a series of unsupervised networks, each of which forms topographic maps. These maps may provide preprocessed input to other cortical modules and to the hippocampal region. We focused on piriform cortex because of its simple paleocortical structure and its direct connection to the hippocampal region. However, there is evidence that similar general principles apply in other sensory cortices (refer to figure 8.3). In fact, topographic mapping may be a general principle of cortical organization. In primary sensory cortex, it is easy to see the topographic representation, because we know which sensory inputs drive individual cortical neurons. In later processing stages, where it is more difficult to know what inputs “mean,” it may be more difficult to observe topographicity. And in cortical areas that are involved with abstract function such as planning or working memory, it is harder even to imagine what a topographic representation might look like. However, researchers are currently investigating these issues as well as other functions of cortex. There are also many computational models that address more complex aspects of cortical operation, including cell synchronization, inhibitory interneurons, and columnar architecture.³⁸

In this chapter, the focus has been on representational plasticity in sensory cortex. The data and models that we have discussed so far deal with unsupervised plasticity, in which cortical representations are changed on the basis of stimulus frequency and the similarities between stimuli (stimulus

clusters). However, there is another important factor governing cortical plasticity: stimulus meaning, such as whether a stimulus predicts future reinforcement (e.g., a US).

Norman Weinberger and colleagues have made extensive study of primary auditory cortex, recording from individual neurons in guinea pigs before and after the animals learn that a tone CS predicts some salient US.³⁹ Initially, each neuron in auditory cortex responds most to a particular tone: This is the cell's **best frequency**; the neuron responds in a decreasing fashion to tones that are increasingly different from this best frequency. For example, figure 8.20A shows the response of a single neuron when various tones are presented. This neuron's best frequency is around 0.75 kHz; the response drops off gradually as the frequency increases, and by 10 kHz, there is little or no response. Similarly, tones below the best frequency also generate weak responding. The best frequency of a neuron is not random, but relates to the neuron's position in the cortical map (refer to figure 8.3B). Thus, nearby neurons presumably have similar best frequencies.

Next, Weinberger and colleagues paired a 2.5-kHz tone CS with a US and observed what happened to the cortical neurons. Some neurons changed their best frequency in the direction of the CS, as is shown in figure 8.20B. That is, cells that previously responded only weakly to the CS now responded more strongly to it. If enough cells showed this type of change, the overall result might be a change in the cortical map: More areas of cortex would respond to the CS (compare figure 8.5). The changes were fast, occurring within as few as five training trials.⁴⁰ As long as eight weeks after training, many neurons maintained their new firing patterns.⁴¹ However, if the CS is presented alone—not with the US—the changes may be absent or greatly reduced.⁴² This implies that *stimulation alone doesn't drive cortical plasticity; instead, the stimulus has to be meaningfully related to reinforcement*.

Apparently, cortical plasticity occurs when a stimulus such as a CS is meaningfully related to (predictive of) a salient reinforcement such as a US. The next obvious question is: How does US information reach primary sensory cortex? USs often take the form of shock or food and are therefore perceived as somatosensory or gustatory stimuli; it is unlikely that this kind of information is accessible to primary auditory cortex, which is devoted to processing acoustic stimuli.⁴³ However, Weinberger's findings clearly indicate that the shock US influences plasticity in auditory cortex.

Weinberger and Merzenich have both suggested the same basic interpretation: They have argued that the auditory cortex does *not* receive information about reinforcing somatosensory or gustatory stimuli. Rather, they believe that it receives information only that a US of some kind has occurred. This

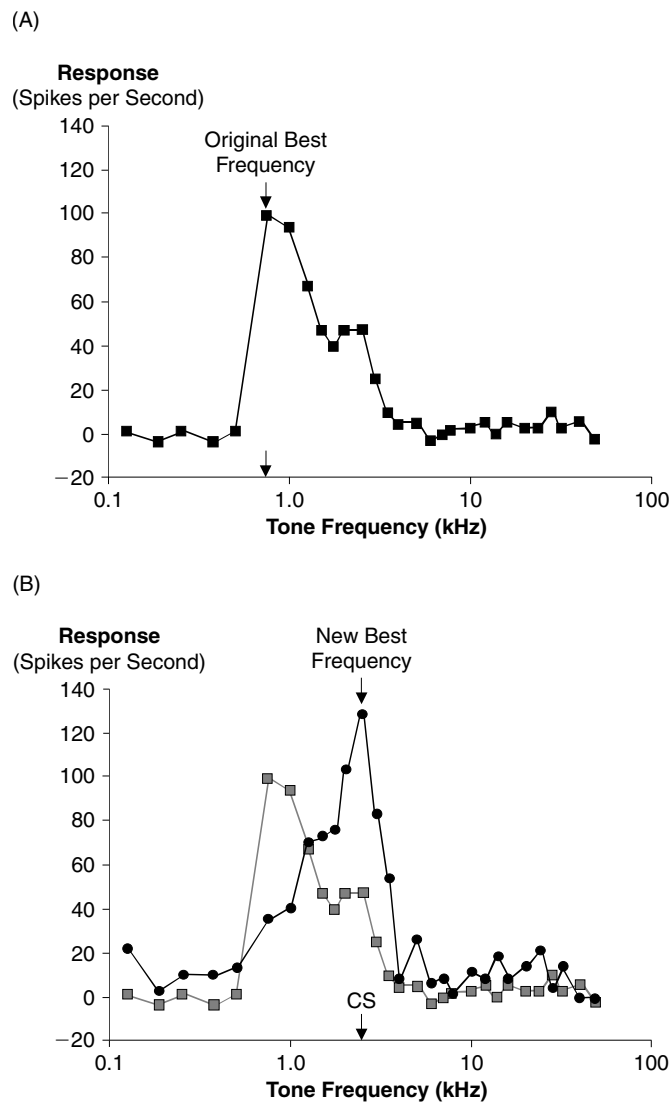


Figure 8.20 Responses of a single neuron in primary auditory cortex of guinea pig to various tone stimuli. (A) Originally, this neuron responds maximally to a tone at frequency 0.75 kHz; this is the cell's best frequency. The response drops off gradually for higher tones and precipitously for lower tones. Other neurons in auditory cortex respond preferentially to other tones. (B) After training that a 2.5-kHz tone CS predicted a shock US, the neuron's responses altered so that the new best frequency was the CS, and there was proportionately less responding to the previous best frequency. (Adapted from Weinberger, 1997, figure 2.)

information is enough to drive cortical remapping to expand the representation of the CS. Other brain areas (such as the amygdala, hippocampus, and polymodal cortex) may be responsible for encoding the memory linking this CS with a particular US—but not primary sensory cortex. Primary sensory cortex knows only which stimuli are worth expanded representation and which are not.

To implement this system, what is needed is a brain region that can signal importance to sensory cortex, giving warning that the current stimulus is worth remapping, without necessarily specifying exactly why. It turns out that several brain regions exist that could serve this kind of a function; both Weinberger and Merzenich have focused on the **nucleus basalis**, a small group of neurons located in an area known as the **basal forebrain** (figure 8.21). The nucleus basalis projects to all areas of the cortex and to the amygdala (though, interestingly, it does not project strongly to hippocampus). When nucleus basalis neurons are activated, they release a neurotransmitter called **acetylcholine** (abbreviated **ACh**). Acetylcholine has many functions in the brain, one of which is to enhance neuronal plasticity. Thus, a simplified view of the function of the nucleus basalis is that it could function as an enabler of cortical plasticity: *When a CS is paired with a US, the nucleus basalis becomes active and delivers acetylcholine to cortex, enabling cortical remapping to enlarge the representation of that CS.*⁴⁴

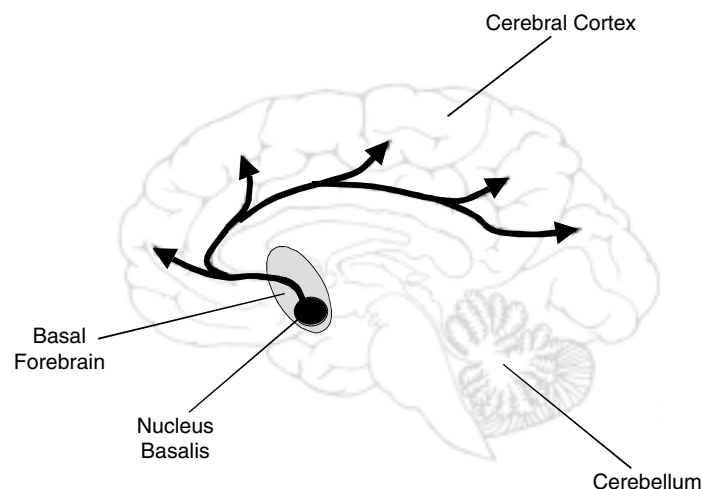


Figure 8.21 Cutaway view of the human brain showing approximate placement of the nucleus basalis within the general area denoted as the basal forebrain. Cholinergic neurons in the nucleus basalis project throughout the cerebral cortex.

Of course, this argument begs the question of how the nucleus basalis knows when to become active, and this is still a topic of intense current interest and research. However, the nucleus basalis does receive connections from areas such as amygdala that are known to receive information about USs such as shock and food.

Several studies have confirmed that the nucleus basalis can play a role in mediating cortical plasticity. Most important, when a tone CS is paired with nucleus basalis stimulation instead of a real US, cortical remapping occurs to that CS.⁴⁵ This lends strength to the argument that sensory cortex does not receive direct US information but only signals from the nucleus basalis that can initiate plasticity. Additionally, when the nucleus basalis is lesioned or the cholinergic projections are otherwise disrupted, cortical plasticity is greatly reduced.⁴⁶

These findings are very exciting, not least because of their implications for rehabilitation after cortical damage. Perhaps it may eventually even be possible to use judicious stimulation of the nucleus basalis to encourage cortical remapping in individuals who have lost the use of one cortical area. Additionally, the idea of cholinergic mediation of cortical plasticity would be relatively simple to instantiate within a computational model. The competitive networks that we discussed in section 8.2 self-organize on the basis of stimulus frequency and stimulus similarity, without regard for stimulus meaning. It would be straightforward to add an additional assumption that plasticity is facilitated or increased in the presence of a cholinergic input that signals that a stimulus is associated with reinforcement.

8.4 IMPLICATIONS FOR HUMAN MEMORY AND MEMORY DISORDERS

Most of the empirical work described in this chapter is based on animal research, but many of the same processes appear to operate in human cortex as well. This section describes one recent project using the principles of cortical remapping observed in monkeys to improve language skills in children.

Most children who are exposed to language throughout their infancy master the incredible complexities of speech generation and comprehension with seemingly little effort. A few children lag behind their peers in the development of language skills, and some of these children develop a cluster of impairments in language and reading that may be diagnosed as dyslexia. Some studies have estimated that 3–10% of all preschoolers have at least some language learning problems that are not attributable to known factors such as hearing impairment, mental retardation, or known brain lesion, and these children develop normally in other ways.⁴⁷

Work by Paula Tallal and colleagues over the past twenty years has identified one subgroup of **language-learning impaired** (LLI) children who have normal intelligence but score well below normal on oral language tests. Tallal tested these children on tasks that required discriminating syllables such as /ba/ and /da/, which begin with consonant sounds that differ only in the initial few milliseconds (see MathBox 8.2). The LLI children were specifically impaired at making these fine temporal discriminations. Thus, Tallal hypothesized that what appeared to be a language-specific deficit in the LLI children might actually reflect a difficulty in processing information that was presented for only a few tens of milliseconds. Consistent with this idea, Tallal found that the LLI children had deficits in a wide range of nonspeech areas that involved such information. For example, if the LLI children were touched in rapid succession on two different fingers, they had trouble identifying which finger was touched first. Thus, LLI may be not a primarily linguistic impairment, but rather a problem in rapid sensory processing that has its most obvious expression in language difficulty.

The next step was to consider how LLI children could be helped. Michael Merzenich and other workers, studying animals, had shown that the cortex was plastic throughout life and that its ability to make distinctions between stimuli could be modified by experience—specifically by intense practice with those stimuli (refer to section 8.1). If that practice required making fine distinctions about temporal differences, then the cortex should remap to emphasize temporal information and facilitate the distinctions. Prior studies had already shown that normal adults who were given intense practice in recognizing extremely brief stimuli could demonstrate great improvements in their performance. Perhaps the same would be true with LLI children.

Tallal and Merzenich devised several computer games that were intended to drill LLI children in acoustic recognition of rapidly presented sounds. One game involved the presentation of two tones, each of which could be either high or low; the children were required to execute keyboard responses depending on the order of tone presentation. Initially, the tones were relatively long (60 msec) and had a brief interval between them (500 msec). The LLI children could master this task easily. As the drill progressed, the stimulus duration and intervals were gradually reduced.

In other games, the children were required to distinguish syllables that had been acoustically modified so that the difficult rapidly changing aspects (such as those differentiating /ba/ from /da/) were artificially extended—for example, from 40 msec to 80 msec. The syllables sound distorted but are easy to discriminate, even for LLI children. Again, the distorted aspects were gradually speeded up until they were as rapid as in normal speech.

Children were drilled on these games several times a day for several weeks, and by the end of this training period, many of the LLI children were able to play the games with the same speed and accuracy as non-LLI children. Following this training, the LLI children were given a test of their ability to process rapidly presented stimuli; performance improved dramatically relative to pretraining levels, and this improvement was still evident at a posttest six weeks after training.⁴⁸ More important, after a summer of practice, many teachers and parents reported improved performance when the LLI children returned to school in the fall.*

This study represents one example of how insights about cortical function discovered in animal research—specifically, that intensive training with particular stimuli could result in cortical remapping that made those stimuli more discriminable—could have clinical implications for humans with learning impairments. Potentially, the same idea could be applied to other domains of human learning impairment.

SUMMARY

- In mammals, cerebral cortex is mainly six-layered neocortex; a few areas such as primary olfactory (piriform) cortex are two-layered allocortex similar to the cortex found in reptiles and birds.
- For each sensory modality, the first cortical processing occurs in a specific region: primary visual cortex, primary auditory cortex, and so on. Primary sensory cortex is usually laid out in a topographic map, in which each subregion of cortex responds preferentially to a particular type of stimulus and neighboring cortical regions respond to similar stimuli.
- Cortical maps are plastic in adults, in response to dramatic changes in stimulation (such as deafferentation) and in response to learning (repeated pairing of a stimulus with reinforcement). Stimulus representations compete for cortical space, so as the area devoted to encoding one stimulus expands, the regions devoted to other nearby stimuli are shifted or compressed.
- Unsupervised or self-organizing networks are a class of neural network that discover underlying regularities in the input without generating any

*Tallal and Merzenich's learning games are commercially available under the product name Fast ForWord. They report that during two years of field testing by 35 speech pathologists and other clinicians, the program had a 90% success rate and helped to improve language processing in over 500 children with specific language impairments. Fast ForWord is marketed by the Scientific Learning Corporation. More information is available at the company's Internet Web site (<http://www.scientificlearning.com>) or by mail from Scientific Learning Corporation, 1995 University Avenue, Suite 400, Berkeley, CA 94704.

predefined output (such as a CR that anticipates the US). Competitive networks are a class of unsupervised network in which nodes compete to respond to inputs. Self-organizing feature maps are a kind of competitive network that can develop topographic representations similar to those developed in sensory cortex.

- Olfactory or piriform cortex may function as a competitive network that performs hierarchical clustering of odor inputs. In the process, it compresses co-occurring odors, meaning that it is harder later to distinguish the individual odor components.
- The cortico-hippocampal model assumes that one function of the hippocampal region is to differentiate the representations of predictive stimuli. This differentiation may ordinarily serve as an opponent process to combat piriform-mediated odor compression when needed. Hippocampal-region damage should then result in a tendency to overcompress co-occurring odors. This is consistent with data showing that rats with hippocampal-region damage have difficulties in simultaneous odor discrimination and in responding appropriately when familiar odors are presented in novel recombinations.
- Cortical plasticity reflects not only stimulus frequency and stimulus similarity but also stimulus meaning. Rather than receiving direct US inputs, sensory cortex receives cholinergic projections from the nucleus basalis that may signal that the current input is important—and representations should be enlarged—without specifying the exact nature of the associated reinforcement.
- Theoretical and empirical findings about cortical plasticity have led to real-world advances, particularly in the treatment of language-learning impairments by intensive training with difficult speech elements.

APPENDIX 8.1 SIMULATION DETAILS

The model presented in the section beginning on page 241 was originally described in Myers & Gluck (1996). Full details of this model and its application to odor discrimination are given there.

The *external input* to the system is a 48-element vector. The first 12 elements represent 12 possible odors; each element is set to 1 if the corresponding odor is present on the current trial or to 0 otherwise. The remaining 36 elements make up a three-element subfield for each of the 12 odors, indicating whether the odor (if present) occurs at the left, center, or right odor port. In the experiments reported in the section beginning on page 241, two odors are presented on each trial, one each at the left and right odor ports.

The *piriform network* is a simplification of the piriform cortex model described in the section beginning on page 232.⁴⁹ It consists of a single layer

of 25 nodes, each receiving weighted connections from all the inputs. There are 156 inputs: the 12 odor inputs each magnified to occupy a ten-element subfield and the 36 spatial inputs. The connection weights and node biases are initialized from the uniform distribution $U[0..1]$ and normalized so that the total weight to each node sums to 1.0. The nodes are divided into five groups of five nodes each. On each trial, each patch individually determines its winning node j with greatest activation V_j determined as $V_j = f(\sum_i w_{ij} V_i + \theta_j)$, where $f(x) = 1/(1 + e^{-x})$, w_{ij} is the weight from input i to node j , V_i is the activation of input i , and θ_j is the bias of node j . The output o_j of each winning node j is set to 1.0, while the outputs of all other nodes in the cluster are set to 0.0. Each node j then updates its weights as $\Delta w_{ij} = \beta(o_j - V_j)V_i$ with learning rate $\beta = 0.005$. All weights are bounded between 0 and 1.

The *hippocampal-region network* is a predictive autoencoder with 61 inputs (the 36 external spatial inputs plus the 25 outputs of the piriform network), 25 internal-layer nodes, and 63 output nodes, which reconstruct the inputs and predict whether the behavioral response was “go left” or “go right.” Weights and biases are initialized from a uniform distribution $U[-0.1..+0.1]$. (Randomly, two lower-layer weights into each internal-layer node are initialized from the uniform distribution $U[-1, +1]$.) Node activations are computed as in the piriform network; weights are trained by error back-propagation as described in MathBox 4.1 with learning rate $\beta = 0.25$ and momentum $\alpha = 0.9$.

The *cortico/cerebellar network* receives the same 61 inputs as the hippocampal-region network and contains 25 internal-layer nodes and 2 output nodes L and R, corresponding to the “go left” and “go right” responses. Weight initialization and node activation are identical to the hippocampal-region network. The behavioral response is computed from the two output nodes according to

$$\Pr(\text{response} = \text{go} - \text{left}) = \frac{1}{1 + e^{10(y_R - y_L)}}$$

The output nodes are trained according to the Widrow-Hoff rule (Math-Box 3.1) with $\beta = 0.5$. On each trial, only the weights from the output node associated with the behavioral response (L or R) are trained; if reward was trained, the desired output for that node was 1.0; otherwise, the desired output was 0.0. The internal-layer nodes are also trained by the Widrow-Hoff rule; the desired output of node j is the output of internal-layer node j in the hippocampal-region network.

The *fornix-lesion model* is identical to the intact model except that there is no learning in the hippocampal-region network; that is, $\beta = \alpha = 0$. As a result, the cortico/cerebellar network adopts hippocampal-region representations which are fixed and arbitrary.

This excerpt from

Gateway to Memory.

Mark A. Gluck and Catherine E. Myers.

© 2000 The MIT Press.

is provided in screen-viewable form for personal use only by members of MIT CogNet.

Unauthorized use or dissemination of this information is expressly forbidden.

If you have any questions about this material, please contact
cognetadmin@cognet.mit.edu.