

# Introduction to Neural and Cognitive Modeling

Second Edition

**Daniel S. Levine**  
*University of Texas at Arlington*



**LEA** LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS  
2000 Mahwah, New Jersey London

decisions on whether to grant mortgage insurance; classification of some types of medical information, both by symptom pattern and by visual display; and classification of radar patterns by emitter. In neurobiology, network models have made some experimental predictions, particularly in vision, and have begun to suggest analyses of more central brain functions. The comfortable interplay between theory and experiment that has existed for most of this century in physics has yet to be firmly established in neurobiology or psychology. But since the early 1990s, an increasing number of neuroscience and psychology laboratories have included computational modeling as part of their operations.

The scientific approach to knowledge, broadly speaking, argues that mental phenomena should have *some* mechanistic basis that will eventually be understandable by human beings. As Wiener (1954, p. 263) said, the faith of scientists is that nature (including mind) is governed by ordered laws, not by the capricious decrees of a tyrant like Lewis Carroll's Red Queen. Neural network modeling provides the best methodology now available for building toward the theoretical neuropsychology of the future.

Since all current neural models are subject to modification, this book is written to give the student or other reader hands-on experience in thinking about, simulating, and ultimately designing neural networks. It begins in Chapter 2 with a historical overview of major trends and the roots of current key ideas. Later chapters are organized around cognitive tasks or structural principles or both. The structures proceed from simpler to more complex interconnected networks as the book proceeds.

## 2

## Historical Outline

*Faithfulness to the truth of history involves far more than a research ... into special facts ... The narrator must seek to imbue himself (sic) with the life and spirit of the time.*

Francis Parkman, *Pioneers of France in the New World*

*The abuse of truth should be as much punished as the introduction of falsehood.*

Blaise Pascal, *Pensées*

### 2.1. DIGITAL APPROACHES

Neural network modeling as we know it today is rooted in a rich interdisciplinary history dating from the early 1940s. Much of this history is discussed in Sections 1 through 4 of Levine (1983) and is reviewed here. The early development of digital computers, and some perceived similarities between computers and brains, spurred interest in developing a new science called cybernetics (Wiener, 1948). In particular, the computer-brain analogy was based on the fact that neurons are all-or-none, either firing or not firing, just as binary switches in a digital computer are either on or off.

Since that time, neurophysiological data have indicated that the all-or-none outlook is oversimplified. Also, neural network models have been developed whose functional units are neuron populations rather than single neurons. In spite of these technical advances, current approaches still owe many of their

formulations to pioneers from the 1940s, such as McCulloch, Pitts, Hebb, and Rashevsky.

### The McCulloch-Pitts Network

This inquiry essentially began with the classical study of all-or-none neurons by McCulloch and Pitts (1943). In this article, hidden under some elaborate symbolic logic, is a demonstration that any logical function can be duplicated by some network of all-or-none neurons. That is, a neuron can be embedded into a network in such a manner as to fire selectively in response to any given spatiotemporal array of firings of other neurons in the network.

The rules governing the excitatory and inhibitory pathways in McCulloch-Pitts networks are the following:

1. All computations are carried out in discrete time intervals.
2. Each neuron<sup>1</sup> obeys a simple form of a *linear threshold law*: it fires whenever at least a given (threshold) number of excitatory pathways, and no inhibitory pathways, impinging on it are active from the previous time period.
3. If a neuron receives a single inhibitory signal from an active neuron, it does not fire.
4. The connections do not change as a function of experience. Thus the network deals with performance but not learning.

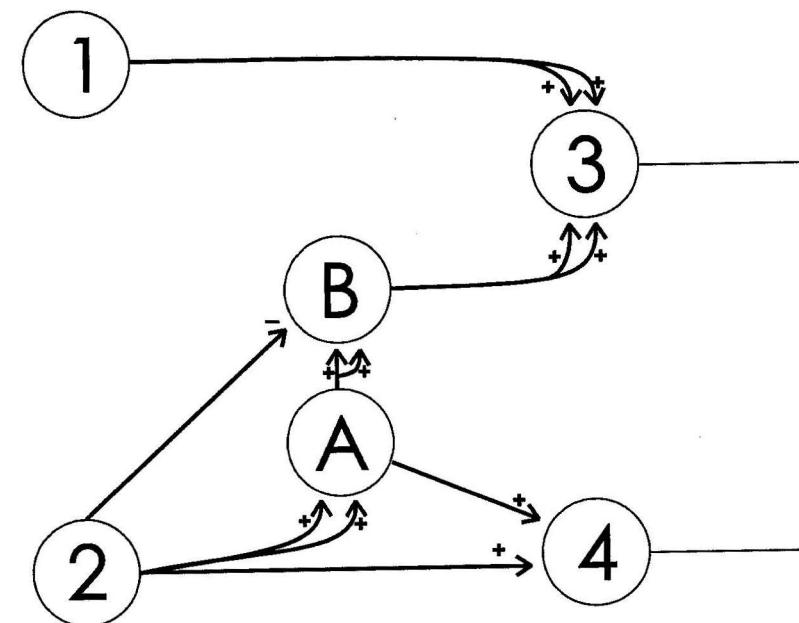
More general linear threshold laws are considered later in this section, in reference to the work of Rosenblatt (1962).

An example of an all-or-none neural network is reproduced in Figure 2.1. This network was designed by McCulloch and Pitts (1943) as a minimal model of the sensation of heat obtained from holding a cold object to the skin and then removing it. The cells labeled "1" and "2" are, respectively, heat and cold receptors on the skin, whereas heat and cold are felt when cells "3" and "4" fire, respectively. Each cell has a threshold of 2, hence fires whenever it receives two excitatory (+) and no inhibitory (-) signals from other cells active at the previous time.

In the network of Figure 2.1, if a cold object is presented and then removed, this means that at Time 1, the only cell firing is cell 2. At Time 2, cell

<sup>1</sup> In most of the models discussed in this book, network elements are called "nodes" or "units" rather than "cells" or "neurons." The exception is made for the McCulloch-Pitts network because their network is directly inspired by the all-or-none firing properties of neurons.

A fires because it receives two excitatory signals from cell 2. Since the cold has been removed, cell 2 does not fire again, nor do any of the other cells in the network. At Time 3, cell B fires because it receives two excitatory signals from cell A. At Time 4, the two excitatory signals from B to 3 cause 3 to fire, meaning that heat is felt. The time sequence of firing patterns is shown in Table 2.1(a). In contrast, consider the same network's response to the cold object being on the skin continuously, as shown in Table 2.1(b). At Time 2, cells 2 and A will both be firing. At Time 3, cell B will not fire because the inhibitory signal from cell 2 prevents B's firing in response to A. Cell 4, however, will fire because it receives excitation from *both* cells 2 and A; hence, cold will be felt.



**Figure 2.1.** Example of an all-or-none network. Neurons labeled "1" and "2" are heat and cold receptors on skin. Heat and cold are felt when neurons "3" and "4" are active, respectively. Each neuron has threshold 2. A cold object held to the skin and then removed causes a sensation of heat. (Adapted from *Mathematical Biosciences*, 66, D. S. Levine, Neural population modeling and psychology: A review, 1-86, Copyright 1983, with permission from Elsevier Science.)

Time	Cell 1	Cell 2	Cell <i>a</i>	Cell <i>b</i>	Cell 3	Cell 4
1	No	Yes	No	No	No	No
2	No	No	Yes	No	No	No
3	No	No	No	Yes	No	No
4	No	No	No	No	Yes	No

FEEL  
HOT  
(a)

Time	Cell 1	Cell 2	Cell <i>a</i>	Cell <i>b</i>	Cell 3	Cell 4
1	No	Yes	No	No	No	No
2	No	Yes	Yes	No	No	No
3	No	Yes	Yes	No	No	Yes

FEEL  
COLD  
(b)

**Table 2.1.** Firings of neurons in the network of Figure 2.1 at successive time steps.

The McCulloch-Pitts model, although it uses an oversimplified formulation of neural electrical activity patterns, presages some issues that are still important in current cognitive models. For example, some of the best known modern connectionist networks contain three types of units or nodes — *input units*, *output units*, and *hidden units*. The input units react to particular data features from the environment (e.g., “cold object on skin,” “black dot in upper left corner,” “loud noise to the right”). The output units generate particular organismic responses (e.g., “I feel cold,” “the pattern is a letter A,” “walk to the right”). The hidden units (a term popularized by Rumelhart & McClelland, 1986) are neither input nor output units themselves but, via network connections, influence output units to respond to prescribed patterns of input unit firings or activities. The input-output-hidden trilogy can at times be seen as analogous to the distinction

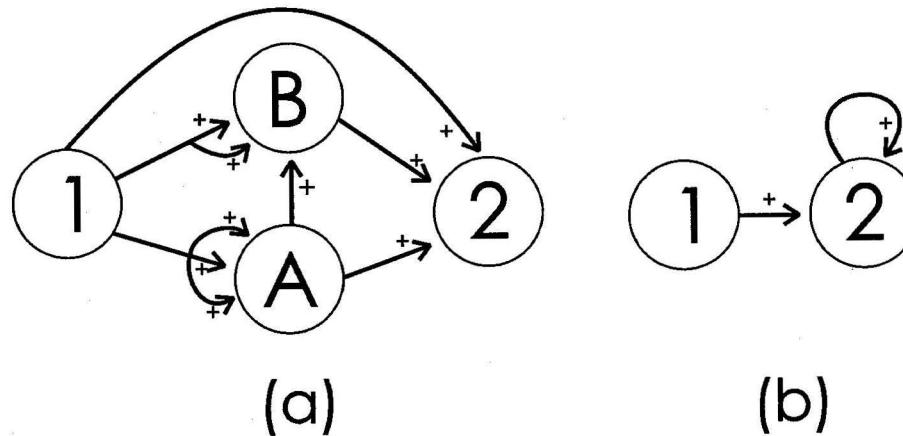
between sensory neurons, motor neurons, and all other neurons (interneurons) in the brain. At other times, though, a model neural network is designed to represent a small part of a larger behavioral process. The output may therefore not be a motor output but a particular internal state, such as a categorization or an emotion, that could be preparatory to a present or future motor response.

Note that in the McCulloch-Pitts network of Figure 2.1, there are already input units (cells 1 and 2), hidden units (cells *A* and *B*), and output units (cells 3 and 4). This distinction becomes explicit in more sophisticated linear threshold networks that are discussed below. In particular, the *perceptrons* developed by Rosenblatt (1962) contained units classified as “sensory,” “associative,” or “response.”

Another cognitive issue raised by the “feel hot when cold is removed” network of Figure 2.1 is how to create output unit responses to given inputs that depend on the context of previous inputs. Specifically, this network responds to difference of the present input from a previous one; this may be called *temporal contrast enhancement*, by analogy with the *spatial contrast enhancement* (particularly observed in visual responses) which is a main topic of Chapter 4. Various forms of temporal contrast enhancement have been combined with learning in many neural network models (e.g., Bear, Cooper, & Ebner, 1987; Grossberg, 1972b, 1972c; Grossberg & Schmajuk, 1987; Klopf, 1986; Sutton & Barto, 1981). Some of these networks model such psychological effects as a motor act becoming rewarding when it turns off an unpleasant stimulus (relief); the withholding of an expected reward being unpleasant (frustration); and the reward value of food being enhanced if the food is unexpected (partial reinforcement acquisition effect).

McCulloch and Pitts also confronted the issue of how memory is stored. Figure 2.2(a) shows a network of the McCulloch-Pitts type in which a neuron fires if a given input (say, a light) is on for three time units in a row. A similar network can easily be constructed to respond to any fixed number of consecutive occurrences of an input. Figure 2.2(b) shows a network in which a neuron is made to fire if the light has been on at *any time in the past*. Note that the mechanism for such memory storage is a reverberatory circuit. The concept of reverberation remains central to the understanding of memory today, and some advantages and limitations of the mechanism are discussed below.

McCulloch and Pitts noted the absence of a precise sense of timing in their model (1943): “the regenerative activity of constituent circles renders reference indefinite as to time past” (p. 130). To them, this makes the model useful in certain ways: “This ignorance, implicit in all our brains, is the counterpart of the abstraction which renders our knowledge useful” (p. 131). Yet obviously a sense of timing is necessary for some other cognitive processes. For those processes, it is necessary to include, as later models do, the possibility of changing connection strengths over time.



**Figure 2.2.** Two more all-or-none neural networks. In both networks, neuron "1" responds to a light being on. (a) Each neuron has threshold 3, and neuron "2" fires after the light has been on for three time units in a row. (b) Neuron "2," which has threshold 1, fires if the light has ever been on in the past.

### Early Approaches to Modeling Learning: Hull and Hebb

At the same time that McCulloch and Pitts were developing a neural network formalism, psychologists were starting to consider mechanistic frameworks for studying learning and memory. This led to consideration of the issue of whether short-term memory (STM) can be distinguished from long-term memory (LTM). Hull (1943) proposed that the two memory processes involved the storage of two sets of traces. For example, consider the classic experiment of Pavlov (1927) where a bell is repeatedly paired with food until a dog salivates to the bell alone. After the experiment is stopped, conscious memory of the bell will be gone, since the dog is concentrating on other things. The memory of the *bell-food association*, however, will still be present, enabling the dog to salivate quickly on the next presentation of the bell. Hull thus distinguished between *stimulus traces* subject to rapid decay and *associative strengths* (or, in his terms, habit strengths) able to persist over a longer time period.

Hull's stimulus traces can be considered as the amounts of activity of particular nodes or functional units in a neural network. His associative strengths,

then, are the strengths of connections between nodes. This suggests first that such connection strengths should change with experience, and second that they should correspond to some variable related to the *synapse*, or junction between neurons.

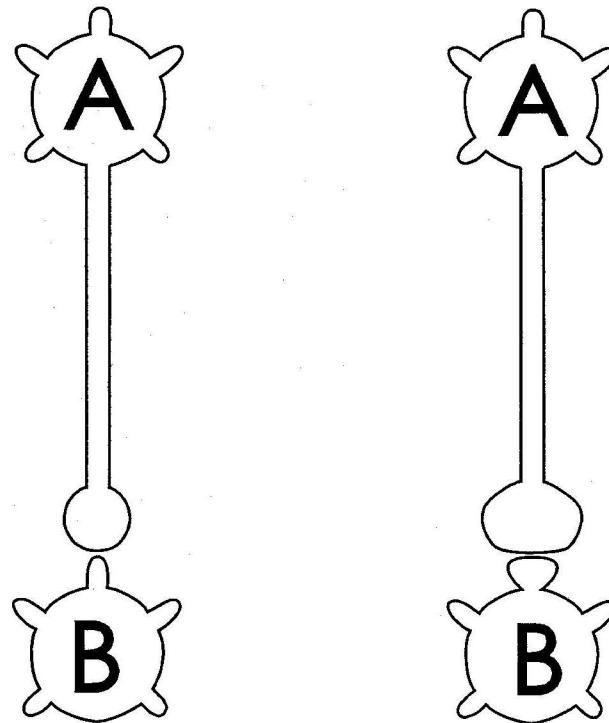
Hebb (1949) interpreted these memory issues with a theory that attempted to bridge psychology and neurophysiology. He declared that reverberatory feedback loops, which had been suggested as a memory mechanism by McCulloch and Pitts (1943), could be a useful mechanism for STM but not for LTM. Concerning traces arising in such reverberatory loops, Hebb (1949) said: "Such a trace would be unstable. A reverberatory activity would be subject to the development of refractory states in the cells of the circuit in which it occurs, and external events could readily interrupt it" (p. 61). He was one of the first to recognize that a stable long-term memory depended on some structural change. But at the same time, he proposed (1949) that "A reverberatory trace might cooperate with the structural change and *carry the memory until the growth change is made*" (p. 62, author's italics).

Hebb went on (1949) to describe a hypothesis for the structural change involved in long term memory: "When the axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased" (p. 62). As for the nature of the structural change, he proposed that if one cell repeatedly assists in firing another, the knobs of the synapse between the cells could grow so as to increase the area of contact (see Figure 2.3).

Neurophysiological data have suggested that actual growth of synaptic knobs can sometimes occur (Anderson et al., 1989; Trommald, Hulleberg, & Anderson, 1996; Tsukahara & Oda, 1981). More frequently, as seen in Chapter 3, there has been experimental support for cellular and synaptic processes that do not involve gross structural changes but that alter the effective strength of connections in other ways. Such processes can embody an associative rule such as Hebb's for changes in connection strength between cells. There has also been extensive theoretical work on alternative rules for learning of connection weights and network modeling based on these rules.

### Rosenblatt's Perceptrons

In the early days of neural network modeling, considerable attention was paid to incorporating Hebb's rule for learning into a network of all-or-none neurons similar to that of McCulloch and Pitts. The modelers building adaptive networks of this variety included Rosenblatt (1962), Widrow (1962), and Selfridge (1959). In these networks, the McCulloch-Pitts form of the linear threshold law was generalized to laws whereby activities of all pathways



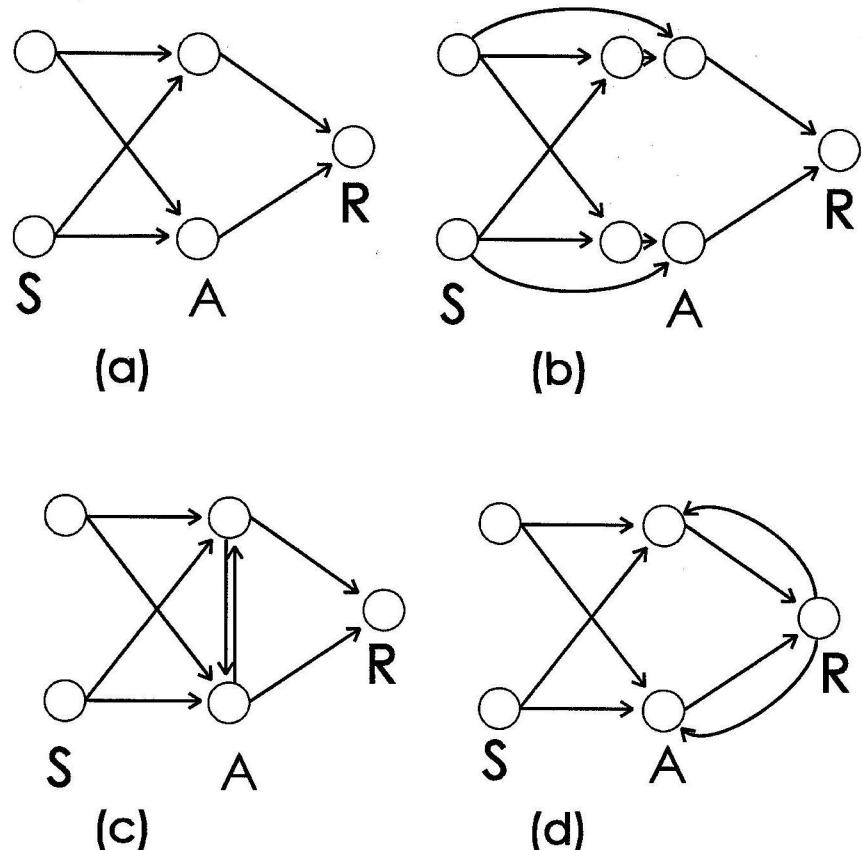
**Figure 2.3.** Diagram of Hebb's structural change hypothesis. The synaptic knob from *presynaptic* cell A to *postsynaptic* cell B gets larger after firing of A is repeatedly followed by firing of B. (Adapted from *Mathematical Biosciences*, 66, D. S. Levine, Neural population modeling and psychology: A review, 1-86, Copyright 1983, with permission from Elsevier Science.)

impinging on a neuron are computed, and the neuron fires whenever some weighted sum of those activities is above a given amount.

The work of Rosenblatt was particularly influential and anticipated many of the themes of modern adaptive networks such as those of the PDP research group (cf. Rumelhart & McClelland, 1986); in fact, the latter type of network are often called *multilayer perceptrons*. The main function he proposed for his perceptrons was to make and learn choices between different patterns of sensory stimuli.

Rosenblatt set out to study the pattern classification capabilities of networks of sensory (*S*), associative (*A*), and response units (*R*) with various structures of

active connections between units. Figure 2.4 shows examples of perceptrons with four possible connection structure types. These types are, in order, *three-layer series-coupled* (connections one-way from *S* to *A* to *R*); *multilayer series-coupled* (connections from *S* to one level of *A* to another level of *A* to *R*); *cross-coupled* (like three-layer series-coupled with the addition of cross links between *A* units, and *back-coupled* (like series-coupled with the addition of feedback links from *R* to *A* units).



**Figure 2.4.** Examples of some classes of perceptrons: (a) three-layer series-coupled; (b) multilayer series-coupled; (c) cross-coupled; (d) back-coupled. (Reprinted from *Mathematical Biosciences*, 66, D. S. Levine, Neural population modeling and psychology: A review, 1-86, Copyright 1983, with permission from Elsevier Science.)

Whatever their type of coupling, Rosenblatt initially considered perceptrons with certain restrictions, which he called *elementary perceptrons*. To define an elementary perceptron, it was first necessary to define a *simple perceptron* (p. 85):

**DEFINITION 22:** A *simple perceptron* is any perceptron satisfying the following five conditions:

1. There is only one *R*-unit, with a connection from every *A*-unit.
2. The perceptron is series-coupled, with connections only from *S*-units to *A*-units, and from *A*-units to the *R*-unit.
3. The values of all sensory to *A*-unit connections are fixed (do not change with time).
4. The transmission time of every connection is either zero or equal to a fixed constant,  $\tau$ .
5. All signal generating functions of *S*, *A*, and *R* units are of the form  $u_i^*(t) = f(\alpha_i(t))$ , where  $\alpha_i(t)$  is the algebraic sum of all input signals arriving simultaneously at the unit  $u_i$ .

**DEFINITION 23:** An *elementary perceptron* is a simple perceptron with simple *R*- and *A*-units, and with transmission functions of the form  $c_{ij}^*(t) = u_i^*(t - \tau) v_{ij}(t)$ .

The precise form of transmission functions and signals is discussed in more detail at the end of this chapter, along with a list of several more of Rosenblatt's mathematical definitions.

Rosenblatt's book consisted of descriptions of a large number of mathematical and computer experiments on how well these different types of networks could either classify or generalize sensory patterns. The approach to modeling was described as *genotypic* rather than *monotypic*. These terms were defined as follows (Rosenblatt, 1962): "Instead of beginning ('monotypic') with a detailed description of functional requirements and designing a specific physical system to satisfy them, this approach ('genotypic') begins with a set of rules for generating a set of physical conditions, and then attempts to analyze their common functional properties" (p. 22).

Finally, we need to consider the learning rules for perceptrons, which Rosenblatt called the *reinforcement system*. Many of his ideas on learning were influenced by those of Hebb (1949). He distinguished two major types of reinforcement systems, *alpha* versus *gamma* systems. In the alpha system, all active connections terminating on a given active cell are changed by equal amounts, whereas inactive connections are not changed at all. In the gamma

system, the total value of connection strengths is conserved, so that inactive connections are decreased while active ones are increased.

The amount of the connection change associated with reinforcement was a value  $\delta$  determined by one of three *training procedures*. In a *response-controlled system*, the magnitude of  $\delta$  is constant and its sign is determined by the response (that is, by the vector of *R*-element activities). In a *stimulus-controlled system*, the magnitude of  $\delta$  is again constant but its sign is determined by the stimulus (that is, by the vector of *S*-element activities). In an *error-correcting system*,  $\delta$  is 0 unless the response is determined elsewhere to be "incorrect." Also, reinforcement can be either *positive* or *negative*, that is, going in either the same direction as or the opposite direction to the current response.<sup>2</sup>

### Some Experiments With Perceptrons

Rosenblatt (1962), starting with Chapter 7, ran experiments in which these different types of perceptrons were taught to discriminate classes of stimuli. A number of distinctions were found between the capabilities of perceptrons with different reinforcement rules and different training procedures, distinctions which are now mainly of historical interest. Not surprisingly, the perceptrons with error-corrective reinforcement converged faster than those with either stimulus-controlled or response-controlled reinforcement. Reinforcement rules of the error-correcting type were concurrently developed by Widrow and Hoff (1960) and are still used widely (e.g., Abdi, Valentin, Edelman, & O'Toole, 1996; Anderson & Murphy, 1986; Cohen & Servan-Schreiber, 1992; Stone, 1986).

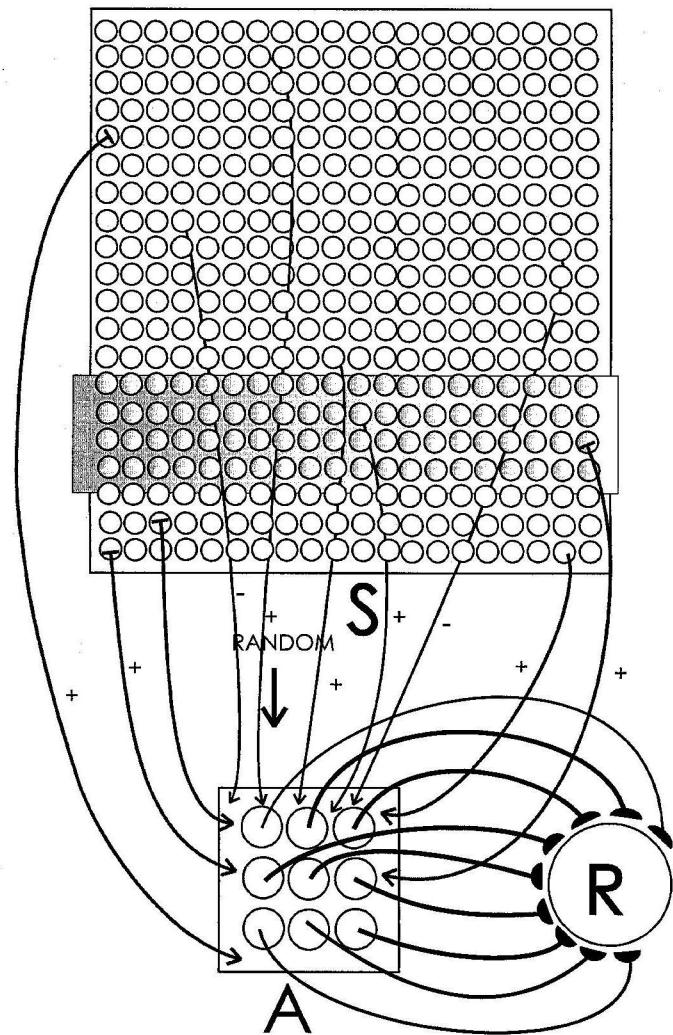
As for the distinction between alpha and gamma reinforcement, the results of the simulation experiments were equivocal. A slight advantage was found for the gamma rule if the various stimuli presented were of unequal size or frequency, whereas the alpha rule seemed to carry some advantage if the system included an error correction mechanism. Conservation laws similar to the gamma rule have been used in more recent neural networks. Rosenblatt found that the conservation rule made the network's responses more likely to be stable. This same property was used in later neural network models by Malsburg (1973) and Wilson (1975), who both thought this "principle of constant synaptic strengths" could be explained in terms of conservation of some chemical substance at or near synapses.

In one of Rosenblatt's major experiments (see Figure 2.5), the *S*-units are arranged in a rectangular grid. Connections from *S*- to *A*-units are random, whereas all *A*-units connect to the single *R*-unit. The perceptron (elementary,

---

<sup>2</sup> This usage differs from the standard usage of experimental psychologists. In psychology, negative reinforcement refers to a stimulus whose removal is rewarding. What Rosenblatt called negative reinforcement, psychologists call punishment.

series-coupled) was taught to discriminate vertical from horizontal bars; variants of this experiment are given in the exercises for this chapter.



**Figure 2.5.** Schematic of a simplified form of one of Rosenblatt's experiments.  $S$  consists of a  $20 \times 20$  grid. Each  $A$  unit receives 5 excitatory and 1 inhibitory inputs from (random)  $S$  units. A horizontal bar input is shaded.

Rosenblatt found that if *all* possible vertical and horizontal bars are presented to the elementary series-coupled perceptron, and the perceptron is reinforced positively for responding to the vertical bars and negatively for responding to the horizontal, then eventually the network gives the desired response reliably to each one. However, if only some of the vertical and horizontal bars are presented and positively or negatively reinforced, the series-coupled perceptron is unable to generalize its behavior to other vertical or horizontal bars that have not been presented. What generalization the network can do is based on location rather than on any more fundamental properties of the input patterns. In models of visual pattern discrimination, issues like translation invariance (ability to recognize a given pattern regardless of where it is in the visual field) remain difficult ones today (see Section 6.5 of this book). This property is exhibited by the Neocognitron of Fukushima (1980) and the What-and-Where filter inspired by architecture of the visual part of the cerebral cortex (Carpenter, Grossberg, & Lesher, 1998).

Inability to generalize is related to another weakness of series-coupled perceptrons: their inability to separate out parts (features) of a complex pattern. This means that for a perceptron to perform categorizations, it needs an excessively large number of nodes. Minsky and Papert (1969) remarked about a similar network that “along with its never-forgetting, it brings other elephantine characteristics” (p. 161). A third weakness is that these systems rely on a reinforcement signal external to the perceptron.

Further experiments and some theorems showed that generalization can be markedly improved by adding more connections to the perceptron. This can be accomplished either by interposing extra layers of associative units or by cross-coupling existing associative units. These additional connections also remove much of the perceptron’s dependence on an external reinforcer. The separation of features from an overall pattern — for example, the classification of patterns on a rectangular grid (“retina”) by whether there is a square in the center — proved more difficult for perceptrons. Preliminary simulations indicated that feature detection might be improved by back-coupling, that is, adding feedback from  $R$  units to  $A$  units.

Rosenblatt developed another idea that is a variant of back coupling — namely, back propagation of errors. This procedure was later mathematically formalized by Werbos (1974; see also Werbos, 1993), modified in various ways by LeCun (1985) and Parker (1985), and then popularized by Rumelhart, Hinton, and Williams (1986). The conceptual principle of back propagation, as stated by Rosenblatt (1962), is: “The procedure to be described here is called the ‘back-propagating error correction procedure’ since it takes its cue from the error of the  $R$ -units, propagating corrections back towards the sensory end of the network if it fails to make a satisfactory correction quickly at the response end” (p. 292). In other words, if some  $A$ -to- $R$  connection strengths need to be corrected

for satisfactory response, inferences can be drawn regarding which *S*-to-*A* connections need to be changed as well.

The mathematics of back propagation, and the possible biological basis for it, are discussed in Chapters 3 and 6. It is one of a variety of neural network schemas related to solving the artificial intelligence problem that Minsky (1961) called *credit assignment*, that is, deciding which part of a system is most responsible for an overall outcome and making the best corrections to the system for changing the outcome in the desired direction. Some more sophisticated credit assignment networks are discussed in Chapter 7, but were presaged by many of the inquiries in Werbos' (1974, 1993) dissertation.

### The Divergence of Artificial Intelligence and Neural Modeling

From the late 1960s to the early 1980s, researchers in artificial intelligence largely abandoned neural networks of the linear threshold variety in favor of heuristic computer programs; this history was discussed in Levine (1983, Section 3.2). During this period, other linear threshold models contemporary with Rosenblatt's had some, although relatively minor, impact on artificial intelligence and neural modeling. Widrow (1962) developed the ADALINE (for "adaptive linear neuron"). Contrary to its author's intentions, this work was more influential among electrical engineers doing signal processing than among any group directly studying intelligent systems (Widrow, 1987). Selfridge (1959) developed the PANDEMONIUM model, which got its name from the different modules called "demons," each of them feature detectors with access to partial information from the environment. Decisions of the entire network were based on a weighted average of the decisions of the different demons. The demon approach had some influence on some early computational models of specific brain areas such as the reticular formation (Kilmer, McCulloch, & Blum, 1969) and the hippocampus (Kilmer & Olinski, 1974). However, at that time, the detailed physiology of these brain areas was not understood well enough for such models to be widely accepted. Selfridge's work also inspired some of the abstract computational geometry of Minsky and Papert (1969).

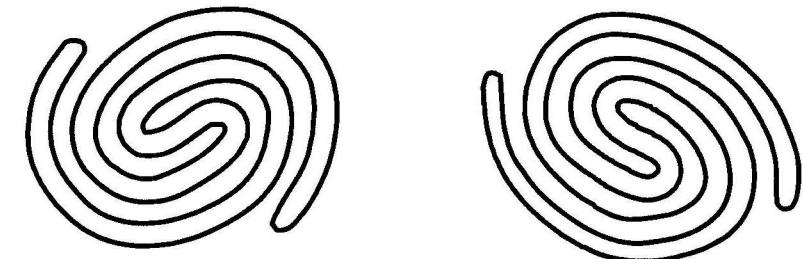
Minsky and Papert (1969) developed their outlook in a book titled *Perceptrons*. The title was inspired by Rosenblatt's previous work, but the devices that Minsky and Papert studied are not exactly a subclass of Rosenblatt's. These abstract machines do, however, have parts that correspond loosely to "sensory," "associative," and "response" areas. The Minsky-Papert perceptron starts with a *retina*, which is a grid consisting of small squares, each of which is at any time active or inactive ("light" or "dark"). Downstream from the retina are units that compute *partial predicates*. Each partial predicate outputs a value of 1 or 0 based on some rule depending on the activity or nonactivity of units in a

given subset of the retina. The maximum size of that subset over all predicates is called the *order* of the perceptron. Finally, there is a decision-making unit that computes a linear function of those predicate outputs and responds when that linear function is above some threshold.

Minsky and Papert proved that their abstract form of the perceptron can learn any classification of patterns on its retina. However, many of the theorems stated that for a perceptron to make some geometrically important classifications, the order of the perceptron has to get arbitrarily large as the size of the retina increases. Theorems of this sort were widely interpreted as discrediting the utility of perceptron-like devices as learning machines. But Minsky later said that, in retrospect, the discrediting of perceptrons seems like an overreaction (Rumelhart & McClelland, 1986, Vol. 1, pp. 158-159).

Moreover, some of the visual discriminations that are difficult for perceptrons are also difficult for humans. For example, consider the distinction between connected and disconnected figures, as shown in Figure 2.6. It is easy for the unaided eye-brain combination to tell that a filled-in circle is connected, whereas a pattern of two filled-in circles side by side is disconnected. But it is next to impossible for the eye and brain to tell which of the two convoluted patterns in Figure 2.6 is connected and which is disconnected without some help from finger tracing.

The models of the group of researchers that called themselves PDP, which originated about 1981 and most of which are summarized in Rumelhart and McClelland (1986), recaptured some of the threads from Rosenblatt's work. They showed that some of the distinctions that are impossible for Minsky and Papert's kind of simple perceptrons (such as between inputs that activate an odd versus an even number of retinal units) can be made by perceptrons with additional "hidden



**Figure 2.6.** The finite-order perceptrons of Minsky and Papert (1969) cannot tell that the curve on the left is connected, whereas the curve on the right consists of two disjoint arcs. Can you tell that by visual inspection? (Reprinted from Minsky and Papert, 1969, with permission of MIT Press.)

unit" layers (cross-connections) *and* nonlinear activation functions. Some of this work is discussed in Chapters 3 and 6.

## 2.2. CONTINUOUS AND RANDOM NET APPROACHES

While the cybernetic revolution was stimulating discrete (digital) models of intelligent behavior, there was a concurrent proliferation of results from both experimental neurophysiology and psychology. Some of these experimental results stimulated the development of continuous (analog) neural models. We turn now to the study of continuous approaches, random net approaches, and finally some partial syntheses of continuous and discrete approaches.

### Rashevsky's Work

One of the pioneers in the development of continuous neural models was Rashevsky. The best exposition of his outlook was in his 1960 book, *Mathematical Biophysics*. The first edition of this book had been written in 1938 — 5 years before the seminal article of McCulloch and Pitts (1943). Subsequently, the evolution of his thinking had been altered by the McCulloch-Pitts article (which was published in a journal that Rashevsky himself founded and edited).

In most applications of mathematics to physical phenomena, including the biophysics of electrical current flow in single neurons, there are variables that are not all-or-none but may take on any of a range of values. Hence, such processes are typically modeled using differential equations, which are equations describing continuous changes over time in an interacting collection of physical variables. (For those desiring a "primer" in differential equations and their utility in neural network modeling, please refer to Appendix 2.) Rashevsky (1960) described how the earlier edition of his book had used differential equations to model various data in the psychophysics of perception. These data included the relation of reaction times to stimulus intensities, and the just noticeable differences among intensities.

Rashevsky went on to describe how his thinking had been influenced by the article of McCulloch and Pitts (1943), which used all-or-none neurons. He stated (Rashevsky, 1960) that "the proper mathematical tool for representing the observed *discontinuous* interaction between neurons was not the differential equation but the Boolean Algebra or Logical Calculus." Yet it was difficult to model the observed psychophysical data using the McCulloch-Pitts postulates. This paradox was resolved with the observation that such behavioral data reflect the combined activity of very large numbers of neurons. Hence, "the

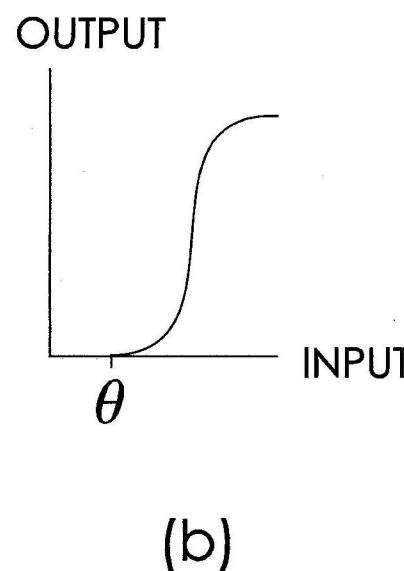
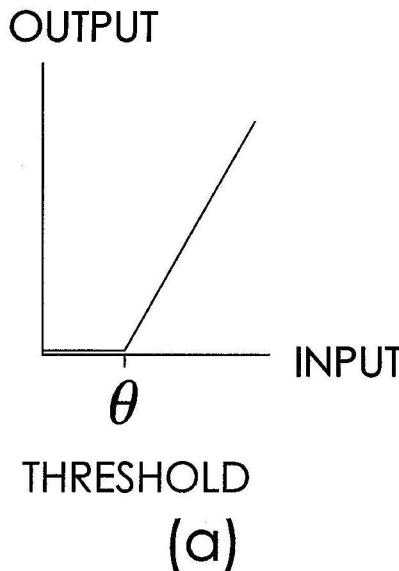
discontinuous laws of interaction of individual neurons lead to a sort of average continuous effect which is described by the differential equations postulated originally" (also from p. 3 of Rashevsky, 1960).

The reconciliation effected by Rashevsky and others between continuous and discrete models is still in common use today. The description in terms of average activity is in line with the trend toward building models based on functional units or nodes that may represent large numbers of neurons (cf. Chapter 1). This is an idea that actually dates back to Hebb (1949), who proposed that significant percepts or concepts are coded not by neurons but by groups of neurons that he called *cell assemblies*.

The boundaries of "functional units" or "cell assemblies" in actual mammalian brains have yet to be defined precisely. Edelman (1987) speculated that units on the order of several thousand neurons in size encode stimulus categories of significance to the animal. Abeles (1991), Burnod (1988), and others have stressed the functional importance of cell assemblies in the mammalian cerebral cortex, or outermost brain layer, that are arranged roughly in columns. Other theorists (e.g., Crick & Koch, 1990; Koch & Crick, 1994; Milner, 1974) have speculated that significant concepts or percepts could be coded by the synchronized electrical activity of large distributed groups of neurons, an idea that has received some neurophysiological support (Eckhorn et al., 1988; Gray, König, Engel, & Singer, 1989; Gray & Singer, 1989).

Whatever its neurobiological mechanism turns out to be, averaging across many neurons also allows the use of deterministic equations for unit activity even if the behavior of single neurons includes a random component. A neuron fires (i.e., transmits an impulse or, more technically, an *action potential*) if its transmembrane voltage exceeds a value called the *threshold* (see Appendix 1 for details). This threshold is widely believed to vary according to some probability distribution, such as the Gaussian or normal distribution (see below). Neural models frequently average such random single-neuron effects across the functional groups of neurons that constitute network nodes; hence, the interactions between nodes become deterministic. In addition, contemporary models average random effects over short time intervals, so that the node activity variable is interpreted as representing a firing frequency rather than a voltage.

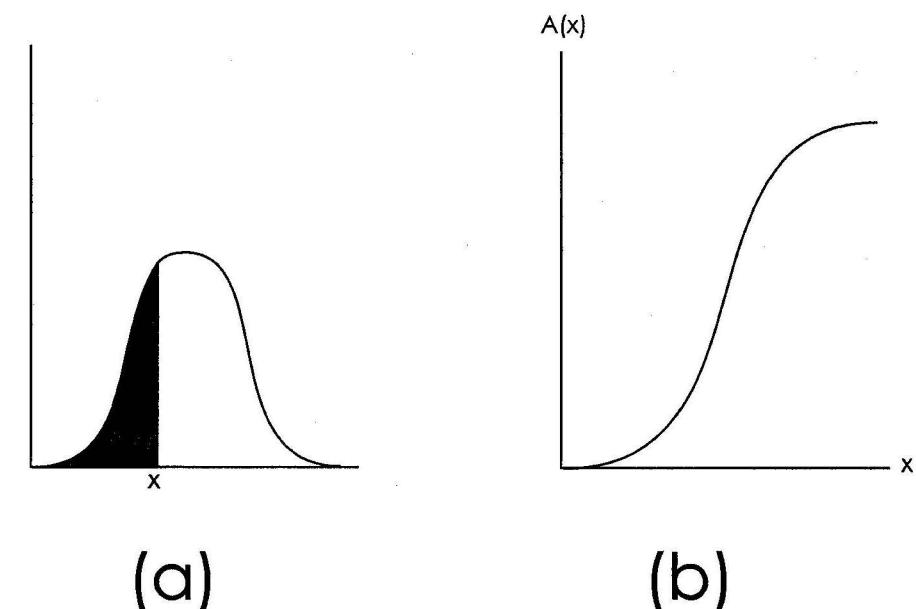
Rashevsky, however, made some simplifying assumptions about the neural averaging process, assumptions that may not always be valid. For example, he assumed that the frequency of impulses transmitted by a neuron is, on the average, a linear function of the cell's suprathreshold activity (see Figure 2.7a). That has proved to be a useful assumption for some neural models of sensory transduction, such as the model of the horseshoe crab retina developed by Hartline and Ratliff (1957). Yet averaging considerations can also lead one to consider input-output functions that are nonlinear, such as *sigmoid* functions (Figure 2.7b). As shown in Figure 2.8, if the firing threshold of an all-or-none neuron is described by a



**Figure 2.7.** Schematic of linear (a) and sigmoid (b) functions of supra-threshold activity.

random variable with a Gaussian (normal) distribution, then the expected value of its output signal is a sigmoid function of activity. For this reason sigmoids have become increasingly popular in recent neural models, as is seen particularly in Chapter 4. Also, there has been some physiological verification of sigmoid input-output functions at the neuron level (Kernell, 1965; Rall, 1955).

For all the weaknesses of his approach, Rashevsky inspired a generation of models that incorporate known neural phenomena into large networks of neurons connected more or less at random. One of these phenomena is the graded (not all-or-none) electrical potentials that occur at the dendrites of a neuron in response to all-or-none action potentials at other cells connected to it. Another is the *refractory period*, the short period of time in which a cell that has just fired (had an action potential) must remain inactive. (A historical outline of some relevant experimental findings appears in Katz, 1966.) Some of these models incorporated the averaging considerations described above, but others used units that were explicitly treated as single neurons. In the terminology of Rosenblatt (1962; see Section 2.1), these random net models tend to be genotypic rather than monotypic.



**Figure 2.8.** One possible biological basis for sigmoid functions: (a) Gaussian (normal) distribution of firing thresholds. If the activity (transmembrane voltage in the case of a single cell) is  $x$ , the node fires if the threshold is less than  $x$ . The probability of that happening is the area under the shaded part of the curve. (b) Schematic graph of the area in  $A(x)$  (a) as a function of  $x$ .

### Early Random Net Models

Many of the early random net models were discussed in the last sections of the review article by Harmon and Lewis (1968).<sup>3</sup> The first attempts at random net modeling include only excitatory connections and no inhibitory ones. The absence of inhibition in model networks, which is unrealistic from the standpoint of known neuroanatomy, also led to unrealistic patterns of electrical activity. The

---

<sup>3</sup> To illustrate the field's rapid growth: Harmon and Lewis (1968) reviewed the literature on *both* neural network and single neuron models in 78 pages. Levine (1983) reviewed network models, but not single neuron models, in 86 pages. This book is taking almost 500 pages to review the network literature in 1998 and its coverage is still incomplete.

excitatory nets developed by Beurle (1956) and Ashby, Foerster, and Walker (1962) tend, as time becomes large, to approach one of two extremes of activity: maximal activity leading to saturation of the entire net, or quiescence. The intermediate level of activity found in actual brains was not modeled by these nets. Griffith (1963a, 1963b, 1965) showed that in this random net framework, stable submaximal activity is possible if inhibition is included.

In the years following Griffith's articles, other modelers tried to develop general theories for random neural networks with both excitatory and inhibitory connections. Most of these theories were based on differential or difference equations that include probabilistic terms. In addition, there were some neural net models inspired by specific formalisms from other scientific fields. Examples are models derived from statistical mechanics (Cowan, 1970) and from nonequilibrium thermodynamics (Freeman, 1975a; Katchalsky, Rowland, & Blumenthal, 1974; Prigogine, 1969). Application to neural network modeling of analogies with other fields continues to this day. Some recent neural networks, for example, have been described as arrays of two-state units. This has led to analogies with the physics of *spin glasses*, which are structures with an array of magnetic spins that have one of two possible values (Amit, Gutfreund, & Sompolinsky, 1985; Chowdhury, 1986; Hopfield, 1982). More examples of physical analogies are discussed in Section 7.1.

Yet analogies are limited by the fact that many nervous system properties are uniquely neural, brainlike, or cognitive. Hence, the further development of continuous and random models since the early 1970s has been influenced less by specific mathematical structures than by neuroanatomical, neurophysiological, and behavioral data. In particular, some data have indicated that brain connections may be random *within* certain neural populations and specific *between* these populations.

### Reconciling Randomness and Specificity

The classic experiments of Lashley (1929) showed that many psychological functions, such as ability to remember specific events, are retained after extensive brain lesions. Lashley's experiments were among the first to inspire the idea, by now common, that representations of events are distributed throughout the brain rather than localized. Other experiments showed, however, that specific connections are important for other functions. Mountcastle (1957) found that the somatosensory (touch-sensitive) area of the cerebral cortex includes a well-organized topographic encoding of the body. Similarly, Hubel and Wiesel (1962, 1965) found that cells in the visual area of the cortex are organized into columns that code specific retinal positions or line orientations. (It is important to note, however, that visual and somatosensory maps are modifiable; the somatosensory maps, at least, can be altered even in adult life.

If the connection to a given area of the cortex from the retinal or body area it would normally code is either cut or inactivated, the same area of cortex can learn to code a different, nearby area. Some of this evidence is summarized in Edelman, 1987).

The paradox between the Lashley data and the Hubel-Wiesel or Mountcastle data is resolved by means of a principle described in Anninos, Beek, Csermely, Harth, and Pertile (1970) as "randomness in the small and structure in the large" (p. 121). This section considers some models whose equations are explicitly based on this principle. The same principle is implicit in many models discussed in later chapters. The latter models use purely deterministic equations at the population level that reflect the averaging over large ensembles of probabilistic effects at the single-cell level.

The article of Anninos et al. (1970) is one of a series of related articles (e.g., Anninos, 1972a, 1972b; Harth, Csermely, Beek, & Lindsay, 1970; Wong & Harth, 1973). In this series of models, neurons are organized with random connectivities into "netlets," and netlets in turn are organized deterministically into larger nets. Evidence for such netlets was found, for example, in the organization of the somatosensory and visual areas of the cortex into functional columns (Hubel & Wiesel, 1962, 1965; Mountcastle, 1957).

Using many cell properties such as refractory periods, Anninos et al. (1970) derived an expression for the expected activity (defined as fractional number of neurons firing) at (discrete) time  $n + 1$  as a function of activity at time  $n$ . The crucial variable for determining long-term behavior is a parameter  $\delta$  describing the number of excitatory postsynaptic potentials (here from within a netlet) needed to cause a cell to fire in the absence of inhibitory inputs. If  $\delta$  is very small, netlet activity always tends to a unique positive stable steady state. If  $\delta$  is very large, netlet activity always tends to 0. If  $\delta$  is in a middle range, there are two stable steady states, one quiescent and one active, and a threshold exists for reaching the active state.

Anninos (1972a) pursued these principles of network organization further with simulations of multinetlet nets. He found, for example, that the dependence of activity of a single netlet in such a network on some external input can exhibit *hysteresis cycles*. That is, the effect of an input can depend on the past history of stimulation. He hinted, without giving details, that such hysteresis could be a mechanism for short-term memory.

Amari (1971, 1972, 1974) described random networks by means of differential or difference equations with two variable parameters — averaged connection weight and averaged threshold. Depending on the values of these two parameters, the network can have either a single stable steady state, many, or none. If the system has excitatory and inhibitory subnetworks, there can be

oscillations of very long period. Amari's systems also modeled association of ideas, by means of connection weights.

A confluence of random net modeling with experimental data occurred in the work of Freeman (1972a, 1972b, 1975a, 1975b), much of which led to models of the olfactory cortex. He laid out some general principles for forming waves from pulses in large neural masses, and showed how this neural mass theory could be used to model EEG (brain wave) patterns and predict their frequencies. He has continued this general line of work to the present, with some results indicating that EEG patterns in the olfactory cortex tend to be chaotic (in the mathematical sense) in the absence of an odorant stimulus, but synchronized in the presence of an odor (Freeman, 1992; Skarda & Freeman, 1987).

Some contemporary modelers are still building networks by connecting neural elements more or less at random and "seeing what happens." The mainstream, however, has shifted from random models to deterministic models that average out the random effects. A mathematical justification for this averaging process, using stochastic differential equations, was given by Geman (1979, 1980). In the deterministic approach, the networks often have particular connection patterns suggested by the cognitive task involved (such as associative learning, pattern storage, selective attention, or categorization). The next four chapters consider approaches to modeling each of those processes, and include discussion of the physiological basis for such models when it is known.

---

### DEFINITIONS AND DETAILED RULES FOR ROSENBLATT'S PERCEPTRONS

Since Rosenblatt's perceptrons are conceptual ancestors of some current popular neural network models, it is useful to study some examples of perceptrons quantitatively. To that end we shall repeat enough of the definitions of his network concepts to be able to formulate exercises for computer simulation. We begin with the definitions of types of units and transmission functions on pages 81 and 82 of Rosenblatt (1962)<sup>4</sup>:

---

<sup>4</sup>The terminology used in this section for unit activities and connection weights is quoted directly from Rosenblatt's book. Hence it differs from the usage in other chapters, which has been made as uniform as possible. Connection weights, labeled  $v_{ij}$  here, are called  $w_{ij}$  in Chapters 3-7. Node activities and signals, denoted by a variety of symbols depending on the type of unit, are usually called  $x_i$  and  $y_j$  in later chapters.

#### Detailed Description: Perceptron to Discriminate Vertical Versus Horizontal

This is a description of the time course of Rosenblatt's simulation of teaching an elementary perceptron to distinguish between 20-by-4 vertical bars and 20-by-4 horizontal bars. As in Figure 2.5, S-units are arranged in a  $20 \times 20$  grid ("retina"). There are 9 A-units and 1 R-unit. Only the A-to-R connections are modifiable.

Each A-unit receives 8 excitatory and 2 inhibitory connections from S-units, and one needs to program a random number generator to find out *which* S-units they come from! The connection strengths  $v_{ij}$  are +1 for excitatory pathways and -1 for inhibitory pathways. For each of the nine A-units, a program similar to the one listed in Figure 2.9 must be used 10 times, one for each connection from an S-unit, to generate a random number uniformly distributed between 0 and 1. Then the random number generated is multiplied by 400 and truncated to an integer, and 1 is added to get an integer between 1 and 400. The last integer determines which S-unit connects to the current A-unit, the S-units in the first row being numbered 1 through 20, the second row 21 through 40, and so forth. More than one connection to that unit can come from the same S-unit. But the location of the S-to-A connections, once set, remains fixed throughout the simulations.

For example, suppose the random number generator applied ten times yields .3171, .0295, .3246, .4878, .9135, .7076, .3168, .5040, .0511, .2607. the locations obtained for the S-unit connections will be 127, 12, 130, 196, 366, 284, 127 (again), and 202 (all excitatory) then 125 and 1 (both inhibitory).

The input stimuli are horizontal or vertical bars of width four. The connectivity within the S grid is *toroidal*: that is, the top row is considered to be adjacent to the bottom row and the leftmost column to the rightmost column. The topmost horizontal bar activates units 1 through 80; the second horizontal bar activates units 21 through 100, and so on down to the twentieth and last horizontal bar, which activates the bottom row and the top three rows of units, that is units 381 through 400 and 1 through 60. Likewise, the leftmost vertical bar activates the left four columns of units, that is, units 1, 21, 41, ..., 381, 2, 22, 42, ..., 382, 3, 23, 43, ..., 383, 4, 24, ..., 384, up through the last vertical bar which activates units 20, 40, ..., 400, then 1, 21, 41, ..., 381. OBJECT: To teach R to respond positively to vertical, negatively to horizontal.

If the  $i^{\text{th}}$  S-unit is activated, then  $s_i^*(t) = 1$ , otherwise  $s_i^*(t) = 0$ . An A-unit computes  $a_j = \sum_i s_i^*(t) v_{ij}$ , and its activity  $a_j^*(t) = 1$  if  $a_j > 2$ , 0 if  $a_j \leq 2$ , 2 being set as the threshold for A-unit activation.

For example, suppose we are looking at the  $i^{\text{th}}$  A-unit, and that the input is the horizontal bar activating rows 5 through 8 of the retina. This means S-units 81 through 160 are activated, so  $s_i^*(t) = 1$  for  $I$  between 81 and 160, and 0 otherwise. The S-to-A connection weight  $v_{ij} = 1$  for  $I = 127$  (twice) and 130 (once) and -1 for  $I = 105$ , because those are locations of excitatory and inhibitory connections. Those being the only S-units for which both  $s_i^*(t)$  and  $v_{ij}$  are nonzero,  $a_j = \sum_i s_i^*(t) v_{ij} = 1(1) + 1(1) + 1(-1) = 2$ , so  $a_j^*(t) = 1$ , that is, that A-unit is activated.

At the start of each run, the  $A$ -to- $R$  connection strengths  $w_j$  are set to values that are randomly (uniformly) distributed between -1 and 1. Error-correction and alpha reinforcement are used. That is, whenever a horizontal line is input and  $R$  (incorrectly) responds positively, any  $w_j$  that are positive while  $a_j^*(t) = 1$  are reduced by an amount  $\delta$ ; whenever a vertical line is input and  $R$  (incorrectly) responds negatively, any  $w_j$  that are negative while  $a_j^*(t) = 1$  are increased by the same value  $\delta$ .  $R$  in turn responds negatively if  $\Phi = \sum_i a_j^*(t) w_j \leq 0$ , and positively if  $\Phi > 0$ . Time delays can all be set to 0.

**DEFINITION 6:** A *sensory unit* (S-unit) is any transducer responding to physical energy (e.g., light, sound, pressure, heat, radio signals, etc.) by emitting a signal which is some function of the input energy. The input signal at time  $t$  to an S-unit  $s_i$  from the environment,  $W$ , is symbolized by  $c_{wi}^*(t)$ . The signal which is generated at time  $t$  is symbolized  $s_i^*(t)$ .

**DEFINITION 7:** A *simple S-unit* is an S-unit which generates an output signal  $s_i^* = +1$  if its input signal,  $c_{wi}^*$  exceeds a given threshold,  $\Theta_i$ , and 0 otherwise.

**DEFINITION 8:** An *association unit* (A-unit) is a signal generating unit (typically a logical decision element) having input

and output connections. An A-unit  $a_j$  responds to the sequence of previous signals  $c_{ij}^*$  received by way of input connections  $c_{ij}$ , by emitting a signal  $a_j^*(t)$ .

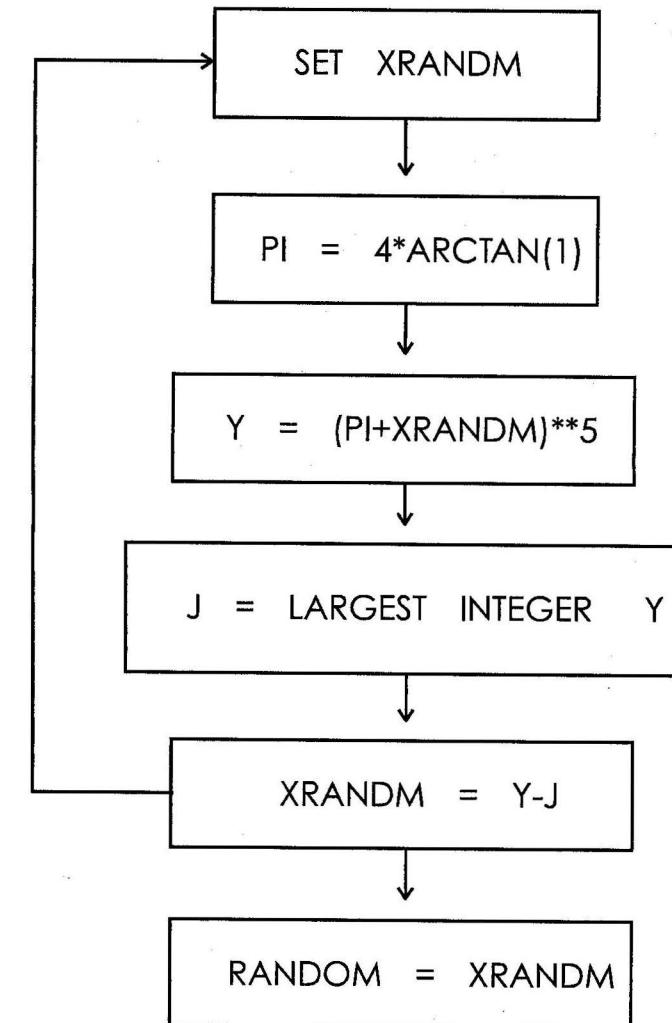


Figure 2.9. Generic program segment that, starting from any initial “seed” number between 0 and 1, generates a different random number in that interval on each pass.

**DEFINITION 9:** A *simple A-unit* is a logical decision element, which generates an output signal if the algebraic sum of its input signals,  $\alpha_i$ , is equal or greater than a threshold quantity,  $\Theta > 0$ . The output signal  $a_i^*$  is equal to +1 if  $\alpha_i \geq \Theta$  and 0 otherwise. If  $a_i^* = +1$ , the unit is said to be *active*.

**DEFINITION 10:** A *response unit (R-unit)* is a signal generating unit having input connections, and emitting a signal which is transmitted outside the network (i.e., to the environment, or external system). The emitted signal from unit  $r_i$  will be symbolized by  $r_i^*$ .

**DEFINITION 11:** A *simple R-unit* is an *R-unit* which emits the output  $r^* = +1$  if the sum of its input signals is strictly positive, and  $r^* = -1$  if the sum of its input signals is strictly negative. If the sum of the inputs is zero, the output can be considered to be equal to zero or indeterminate.

**DEFINITION 12:** *Transmission functions* of connections in a perceptron depend on two parameters: the *transmission time* of the connection,  $\tau_{ij}$ , and the *coupling coefficient* or *value* of the connection,  $v_{ij}$ . The transmission function of a connection  $c_{ij}$  from  $u_i$  to  $u_j$  is of the form:  $c_{ij}^*(t) = f[v_{ij}(t), u_i^*(t-\tau_{ij})]$ . Values may be *fixed* or *variable* (depending on time). In the latter case, the value is a *memory function*.

The concepts relating to reinforcement are defined precisely in pages 88-92 of Rosenblatt (1962), as follows:

**DEFINITION 33:** *Positive reinforcement* is a reinforcement process in which a connection from an active unit  $u_i$  which terminates on a unit  $u_j$  has a value changed by a quantity  $\Delta v_{ij}(t)$  (or at a rate  $dv_{ij}/dt$ ) which agrees in sign with the signal  $u_j^*(t)$ .

**DEFINITION 34:** *Negative reinforcement* is a reinforcement process in which a connection from an active unit  $u_i$  which

terminates on a unit  $u_j$  has its value changed by a quantity  $\Delta v_{ij}(t)$  (or at a rate  $dv_{ij}/dt$ ) which is opposite in sign from  $u_j^*(t)$ .

(Note: the “active units”  $u_i$  in the above definitions could be either *A-units* or *R-units*.)

**DEFINITION 37:** *Alpha system reinforcement* is a reinforcement system in which all active connections  $c_{ij}$  which terminate on some unit  $u_j$  (i.e., connections for which  $u_j^*(t-\tau) \neq 0$ ) are changed by an equal quantity  $\Delta v_{ij}(t) = \delta$  or at a constant rate while reinforcement is applied, and inactive connections ( $u_j^*(t-\tau) = 0$ ) are unchanged at time  $t$ .

**DEFINITION 38:** *Gamma system reinforcement* is a rule for changing the values of the input connections to some unit, whereby all active connections are first changed by an equal quantity, and the total quantity added to values of the active connections is then subtracted from the entire set of input connections, being divided equally among them. Such a system is said to be *conservative in the values*, since the total of all values can neither increase nor decrease. The change in  $v_{ij}$  is equal to

$$\Delta v_{ij}(t) = \left[ w_{ij}(t) - \frac{\sum_i w_{ij}(t)}{N_j} \right] \delta,$$

where  $w_{ij}(t) = \begin{cases} 1 & \text{if } u_i^*(t-\tau) \neq 0 \\ 0 & \text{otherwise,} \end{cases}$

$N_j$  = number of connections terminating on  $u_j$

$\delta$  = reinforcement quantity (typically 1, -1, or 0).

**DEFINITION 39:** A *response-controlled reinforcement system (R-controlled system)* is a training procedure in which the magnitude of  $\delta$  is constant, and the sign of  $\delta$  is entirely determined by the current response,  $r^*$ , regardless of the current stimulus,  $S$ . In general, unless otherwise specified, this term implies that the reinforcement is always positive (i.e., the sign of  $\delta$  agrees with the sign of  $r^*$ , in a simple perceptron).

**DEFINITION 40:** A *stimulus-controlled reinforcement system* (*S*-system) is a training procedure in which the magnitude of  $\delta$  is constant, and the sign of  $\delta$  is determined entirely by the current stimulus,  $S$ , and a predetermined classification ...

**DEFINITION 41:** An *error-correcting reinforcement system* (error correction system) is a training procedure in which the magnitude of  $\delta$  is 0 unless the current response of the perceptron is wrong, in which case, the sign of  $\delta$  is determined by the sign of the error. In this system, reinforcement is 0 for a correct response, and *negative* (see Definition 34) for an incorrect response ... .

## EXERCISES FOR CHAPTER 2

- 1. Hebb's rule for synaptic modification states that a connection strength will increase if activities of the two connected units are both high at the same time (with suitable delays, perhaps). Give some possible advantages and disadvantages of this rule for network models of learning. You may also suggest modifications of this rule, or ways it should be combined with other learning rules.
- 2. Design a McCulloch-Pitts network with heat and cold receptors and a cell that fires after the sequence "heat cold" or the sequence "cold heat" but nothing else. Assume that each cell takes exactly one time step to compute its output, and that cold and heat cannot be simultaneously felt at the same time step. Design another McCulloch-Pitts network, possibly a modification of the first one, so that the last cell fires after alternating sequences of three — "heat cold heat," "cold heat cold," "heat cold neither," or "cold heat neither" — but not after sequences that include repeats — "heat cold cold" or "cold heat heat."
- 3. Do 6 runs of the Rosenblatt elementary perceptron that learns to respond positively to 20-by-4 vertical bars and negatively to 20-by-4 horizontal bars, as described in the "box" on pages 33-34. Do 2 runs with each of three different  $\delta$  values (.20, .5, and .1). Present each of the horizontal and vertical bars, in turn, in any order, repeatedly. It will probably eventually learn to classify all of them correctly but due to the randomness it is not guaranteed to happen. But see how the learning rate depends on  $\delta$ . (Again, because of the randomness,

there is no guarantee about the outcome. But study the tradeoff between the effects of too small or too large a learning rate.)

4. Do the same simulation as in Exercise 3 but with 300 associative units, 3 excitatory connections and 1 inhibitory connection to each *A*-unit, and a threshold of 2.
5. Do the same simulation as in Exercise 4 but teach the network only half of the bars and then present one it has not learned. No generalization should occur.
6. Do the same simulation as in Exercise 4 but with gamma instead of alpha reinforcement.
7. Design a simulation in which a perceptron is trained to discriminate between two types of figures. Examples would be a square of a given size versus a triangle of a given size (that is, translates of a fixed square and a fixed triangle anywhere along the grid) or a square and a diamond. Another example would be to discriminate whether a figure does or does not contain the letter "X."
8. Do a simulation of the ADALINE network of Widrow (1962). In the ADALINE model, a set of bipolar (1 or -1) inputs is filtered through a corresponding set of adaptive weights, and the sum of the weighted inputs is then compared with a desired output. Then error-correcting reinforcement is applied.

The ADALINE equations are as follows. Let  $I_i, I = 1, \dots, n$  represent specific inputs, and  $I_0$  a constant ("bias") input equal to one. Let  $w_i, I = 0, \dots, n$  represent the corresponding weights. Then the actual output, called  $y$ , is 1 if the weighted input

$$S = \sum_{i=0}^n I_i w_i \geq 0$$

and -1 if  $S < 0$ . Let  $y_0$  be a desired output; for example, if the network is trained to learn the logical "AND" operation,  $n = 2$ , and the inputs are 1 and 1, the desired output is also 1. Then at each time step, weights are updated according to the rule

$$\Delta w_i = a(y_0 - y) \frac{I_i}{n+1}$$

Weights are changed until the network has learned the desired output to every bipolar input vector.<sup>5</sup>

For the following simulations, the number  $n$  of nonbias inputs is 2, and the learning rate  $a$  is 1.

- (a) Teach the network to learn the logical “AND” operation, which maps  $(1, 1)$  to 1,  $(1, -1)$  to  $-1$ ,  $(-1, 1)$  to  $-1$ , and  $(-1, -1)$  to  $-1$ . Show that this can be learned in two passes through the sequence of four input vectors.
- (b) Teach the network the logical “OR,” which maps  $(1, 1)$  to 1,  $(1, -1)$  to 1,  $(-1, 1)$  to 1, and  $(-1, -1)$  to  $-1$ . Show that this can be learned in two passes through the sequence.
- (c) Teach the network the logical “NAND,” which always gives the sign opposite to the one given by the “AND.” Show that this can be learned in three passes through the sequence.
- (d) Show that the network *cannot* learn the “exclusive OR,” which maps  $(1, 1)$  to  $-1$ ,  $(1, -1)$  to 1,  $(-1, 1)$  to 1, and  $(-1, -1)$  to  $-1$ , by going through the sequence of training inputs and getting an infinite loop. (The exclusive OR can be learned by multilayer nonlinear networks, as will be seen in Chapter 6).

### Some Additional Sources

**Early neural network models (original articles or reviews):** Anderson, Pellionisz, and Rosenfeld, 1990; Anderson and Rosenfeld, 1988; Palm, 1982; Sejnowski, 1976; Steinbuch, 1961, 1990.

---

<sup>5</sup> Faster learning occurs in the ADALINE if  $y_0$  is compared to the weighted input  $S$  instead of to the quantized output  $y$ . Widrow, Pierce, and Angell (1961) discuss the differences between these two error criteria; in that article, the error we use is called the *neuron error* and the other is called the *measured error*.

# 3

## Associative Learning and Synaptic Plasticity

*The present contains nothing more than the past, and what is found in the effect was already in the cause.*

Henri Bergson, *L'Évolution Créatrice*

*The mind is slow in unlearning what it has been long in learning.*

Seneca, *Troades*

### 3.1. PHYSIOLOGICAL BASES FOR LEARNING

Recall from Section 2.1 the contribution of Hebb (1949) to the bridging of psychology and neurophysiology. Hebb proposed on psychological grounds the existence of synaptic modifications during learning, in the absence, then, of any physiological evidence for such modifications. Since that time, it has been experimentally demonstrated that correlated activity at the pre- and postsynaptic cells of many synapses in animal nervous systems alters the efficacy of the synapse in causing action potentials at the postsynaptic cell. Whereas the relationship of any of these cellular changes to actual storage of cognitive information remains speculative (see Bliss & Collingridge, 1993, or Brown,