# PARALLEL MODELS OF ASSOCIATIVE MEMORY

*Edited by*

## Geoffrey E. Hinton
*M. R. C. Applied Psychology Unit*
*Cambridge, England*

## James A. Anderson
*Brown University*

**EA** LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1981   HILLSDALE, NEW JERSEY

---

## 1

# Models of Information
# Processing in the Brain

James A. Anderson
*Brown University*
Geoffrey E. Hinton
*M.R.C. Applied Psychology Unit*
*Cambridge, England*

### 1.1.  INTRODUCTION

This chapter introduces some models of how information may be represented and processed in a highly parallel computer like the brain. Despite the staggering amount of information available about the physiology and anatomy of the brain, very little is really known about the nature of the higher-level processing performed by the nervous system. There is no established theory about the kinds of neural activity that occur when we hear a sentence, perceive an object, or form a plan, though data on many fascinating and significant bits and pieces is now available.

An obvious feature of the brain is its parallelism (see Section 1.5 for a review of the neurophysiological evidence). This parallelism is a major reason for investigating computational models other than the conventional serial digital computer in our attempts to understand how the brain processes information. The concept of parallelism may need some explanation. A system which is parallel at one level of description may well be serial at a higher level. At the level of individual motor neurons, for example, the human motor system is highly parallel. The simultaneous actions of many muscles are necessary for coordinated movement. If, however, the pattern of activity of the whole set of motor neurons is used as a unit of description, the system is strictly serial because only one pattern can exist at a time. Similarly, in a conventional digital computer many electrical events occur in parallel when each machine instruction is executed, but the instructions, considered as units, are executed sequentially. The transition between the parallel and serial levels of description thus occurs at the level of the individual machine instructions.

If many computational operations are performed at once, a system can obviously operate faster. However, one part does not know what the other parts are currently doing because the parts operate simultaneously. This causes serious problems of coordination and lateral information transfer from one part of a parallel system to another. These problems have made it hard to program general purpose computers that execute many instructions at once though numerous special purpose systems have been developed for specific tasks.

We feel that problems of coordination and lateral information transfer are not merely irritating; they are fundamental. They determine the kinds of operations that are easy to implement at the level at which a machine is parallel. These may be much richer than the rather restricted set of primitive operations of a conventional digital computer. For example, the state of activity of a large set of feature-detecting units can determine the state of activity of another large set of units in a single step in a parallel machine.

The idea that a parallel machine may have a different and richer and much more powerful set of primitive operations constrasts sharply with the idea that parallelism should be added on top of existing programming techniques by providing message-passing facilities that allow communication between multiple processors, each of which is a fully fledged conventional computer. The latter approach is obviously a sensible way of extending existing computational techniques, and it is currently under investigation within computer science, but it takes for granted the primitive operations of a conventional digital computer, which are probably an inappropriate computational metaphor for the brain.

As an example of a task for which a conventional computer seems inappropriate, consider the problem of recalling an item when given a partial description of its properties or a description of its relationships to several other items (Norman & Bobrow, 1979). This appears to be a fairly basic human ability. If the partial description is sufficient to identify an item uniquely, the item often just "comes to mind," with no awareness of any deliberate searching. It is relatively easy to implement this kind of access to items in memory if all the partial descriptions that might be used for access are known in advance. However, human memory does not seem to require this. We can access items from partial descriptions that have not been anticipated. This kind of memory, in which the partial contents of an item can be used to retrieve the remaining contents, is called content-addressable memory. It is a desirable thing to have, but it is very hard to implement in a conventional digital computer (a von Neumann machine). The reason for the difficulty is that the von Neumann machine accesses items in memory by using their addresses (locations in memory), and it is hard to discover the address of an item from an arbitrary subset of its contents. As we shall see, if we abandon the idea that the basic method of retrieving items is via their addresses, we can use parallel computation in systems of interconnected simple elements to achieve content-addressable memory.

Von Neumann machines are based on the idea of a sequential central processor operating on the contents of a passive memory in which data-structures simply wait around to be inspected or manipulated. This conception of memory is shared by most psychologists and is embodied in the spatial metaphors we use for talking about the process of remembering. We think of memory as if it were a filing cabinet or warehouse, and the act of recalling an item is referred to as finding it in memory as if each item were in a specific place and could be found only by going to that place. How else could it be?

The memory models presented in this volume assume a very different basic architecture. Instead of a sequential central processor and a passive memory there is a large set of interconnected, relatively simple processors, which interact with one another in parallel via their own specific hardware connections. Changes in the contents of memory are made by forming new connections or changing the strengths of existing ones. This overcomes a major bottleneck in von Neumann machines, which is that data-structures or programs in memory can only have effects via the sequential central processor, so that it is impossible to mobilize a large quantity of knowledge simultaneously.

A consequence of replacing passive memory by simultaneously interacting units is that the addressing mechanism is replaced by specific hardware connections. The addressing mechanism allows the central processor of a von Neumann machine to access any piece of data, provided the address is known. It thereby allows complex data-structures to be stored in memory by simply making one piece of a data-structure contain the address of the next piece. If one piece contains several addresses, branching structures like trees can easily be stored. Such structures appear to be essential for the implementation of complex representations and computational procedures.

Feldman (Chapter 2, this volume) and Fahlman (Chapter 5, this volume) propose that addresses be replaced by specific hardware connections. Some of the other models in this volume also replace addresses by hardware connections but in a less direct manner. They do not replace a single address by a single hardware connection because they do not use the individual processing units to correspond to items in memory. Instead, items correspond to patterns of activity distributed over many simple hardware units, and the ability of an address to link one item to another is implemented by modifying the strengths of many different hardware connections in such a way that the pattern of activity corresponding to one item can cause the pattern corresponding to the other item (see Section 1.2.3 for details).

The idea that a pattern of activity could represent an item requires some explanation. We use the term *distributed representation* to refer to this way of coding information. Although the concepts of distributed representation and parallelism are quite different, distributed representation appears to be a particularly appropriate method of coding for a highly parallel machine.

Suppose we wish to build a system that can recognize any one of a number of items. One approach to this problem would be to have one internal unit that would respond when and only when its particular item occurred. An alternative approach would be to have each internal unit respond to many of the possible input items. Provided only one item is presented at a time, it will be represented by the pattern of activity of the internal units even though no individual unit uniquely specifies the input item. Thus a pattern of activity becomes the basic representation of the item. There is no necessary loss of precision or ability to discriminate; it is just that internal operations are now performed in a different way. Instead of a single unit causing particular effects on other internal representations or on motor output the pattern of activity of many units causes those effects. It is unnecessary to have a separate higher-level unit that detects the pattern of activity and causes the appropriate effects.

## 1.2. SYSTEMS OF SIMPLE UNITS WITH MODIFIABLE INTERCONNECTIONS

This section describes some models in which changes in the strengths of the interconnections in a system of simple units are used to implement category formation and associative memory. Before introducing these models, however, we outline the ideas about "formal" neurons that were largely responsible for the choice of the particular kind of simple unit used in these models.

### 1.2.1. The McCulloch–Pitts Neuron

Probably the best known, and arguably the most influential model of the nervous system, even today, is that proposed in 1943 by Warren McCulloch and Walter Pitts. They approximated the brain as a set of binary elements—abstract neurons which were either on or off—that realized the statements of formal logic. To quote the first sentence of the abstract of their paper (McCulloch & Pitts, 1943):

Because of the "all-or-none" character of nervous activity, neural events and the relations between them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms . . . and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes [p. 115].

One finds in their paper much of the machinery familiar to those who study automata theory: binary elements, threshold logic, and quantized time where the state of the system at the (n + 1)th moment reflects the states of the inputs to the elements at the nth moment. The primary result of their paper was that nets of such neurons were perfectly general in that they could realize any finite logical expression.

This model obviously has practical implications: put together such neurons and you can make a powerful, general computing device. At about the time of the 1943 paper, exactly such a project was underway at the Moore School of Engineering of the University of Pennsylvania. This paper on brain modeling had an influence on John von Neumann when he sketched the logical outline of the first modern digital computer—the first machine with a program stored with the data. In a famous technical report, von Neumann (1945) said:

Every digital computing device contains certain relay like *elements* with discrete equilibria. Such an element has two or more distinct states in which it can exist indefinitely. . . . The relay action manifests itself in the emission of stimuli by the element whenever it has itself received a stimulus of the type indicated. . . . It is worth mentioning that the neurons of the higher animals are definitely elements in the above sense. . . . Following W. Pitts and W. S. McCulloch . . . we ignore the more complicated aspects of neuron functioning . . . [p. 360].

### 1.2.2. Perceptrons

The perceptron, originally developed by Rosenblatt, and related models such as MADALINE and ADALINE developed by Widrow were intensively studied in the early 1960s. These models have now become part of the lore of pattern recognition, and good short introductions are available in many books on pattern recognition, as well as in the classic books, *Learning Machines* (Nilsson, 1965) and *Perceptrons* (Minsky & Papert, 1969).

The basic element in these devices is the threshold logic unit (TLU), which is a particular type of McCulloch-Pitts neuron. The TLU has a number of inputs, say n, each associated with a real-valued weight that plays a role analagous to the "synaptic strength" of inputs to a neuron. The total input to the TLU is an n-dimensional vector, a pattern of activity on its individual input lines. Each component of the input vector is multiplied by the weight associated with that input line and all these products are summed. The unit gives an output of 1 if this sum exceeds its threshold. Otherwise it gives an output of 0. More formally, the output is the truth value of the expression

$$\sum_i f_i w_i > \theta \qquad\qquad (1\text{-}1)$$

where $f_i$ is the activity on the $i$th input line and $w_i$ is its weight, and $\theta$ is the threshold.

A TLU divides the n-dimensional space of possible input vectors into two regions, separated by a hyperplane, one region being associated with an output of 1 and the other with an output of 0. The values of the weights determine the orientation and position of the hyperplane.

*The Perceptron Convergence Procedure.* Threshold logic units were advanced as adaptive pattern recognition devices (Rosenblatt, 1961; Nilsson,

1965). In the standard perception scheme each input line to the TLU is the output of a feature detector that responds to the presence of some feature in an input array. A perceptron can be made to discriminate instances of a particular class of pattern in the input array by associating appropriate weights with the feature detectors, provided any set of appropriate weights exists.

The main reason for the interest in perceptrons during the 1960s was the existence of an automatic procedure for finding a set of weights that would cause the perceptron to respond if and only if a pattern of a particular type was present in the input array. The procedure, known as the perceptron convergence procedure, works by adjusting the existing set of weights whenever those weights would cause the perceptron to give the wrong response to the current input array. If the perceptron would respond with a 1 and the pattern is actually absent, then all the weights of the active features are reduced equally, and the threshold is raised by the same amount. If the perceptron would respond with a 0 when the pattern is actually present, then all the weights of the active features are raised and the threshold is reduced. Some freedom is allowed in the magnitude of the alterations. One strategy is to make all alterations be of a small constant size. Another method is to make the alterations on each trial just large enough to ensure that the perceptron responds correctly to the current input pattern. If the feature detectors have real-valued levels of activity rather than just being on or off, the alterations in the weights must be proportional to the activity levels of the corresponding feature detectors. Proofs that the perceptron convergence procedure works, and precise statements of its conditions can be found in Nilsson (1965) and Minsky and Papert (1969).

*The Limitations of Perceptrons.* The perceptron convergence procedure cannot be applied to devices in which there is more than one layer of modifiable weights between the input array and the decision unit. The reason for this restriction is that there is no good way of deciding which layer of weights ought to be changed when a multilayered device makes an error. It is clear that devices with multiple layers of modifiable weights are more like the brain and are capable of more sophisticated discriminations, but they lack the automatic learning procedure which is the most important characteristic of the simple perceptron.

Certain limitations of the perceptron were apparent almost from the outset. It is incapable of performing the basic logical operation of exclusive-or. There is no way of setting the weights and the threshold of a perceptron with two inputs so that it will respond positively in just those cases where exactly one input is active. Despite this strong restriction, many people thought that perceptrons were a promising model of perception until Minsky and Papert (1969) produced a rigorous analysis of the limitations of perceptions as pattern recognition devices.

One psychologically unrealistic aspect of the perceptron convergence procedure is that no learning occurs when the response is correct. Satisfactory psychological learning models incorporate what might be called ''positive''

learning; that is, learning when the organism is correct usually appears to be more important than learning when a mistake is made. Extensive experimental evidence supports this claim. Statistical learning theory behaves this way (see the collection of papers in Neimark & Estes, 1967). In the simpler versions of statistical learning theory, learning involves only having a correct response *increase* the probability of making that response in the future.

*The Future of Perceptrons.* Human perception is an extremely complex activity involving multiple interacting representations at many levels. A simple perceptron is clearly an inadequate model. A similar but more complex device involving many layers of perceptrons and collateral and recurrent connections may be capable of perception, but there is no known procedure for learning the weights. A common conclusion has been that devices like perceptrons are not worth studying. This is may be correct if perceptrons are considered primarily as learning devices, but there is an alternative approach that has received little attention until recently (Hinton, Ch. 6, this volume; Minsky, 1977, 1980). This new approach is to consider how computation might be organized within a device consisting of many interconnected perceptronlike units. The emphasis is on programming and on ways of representing knowledge and implementing procedures rather than on finding a magic formula that will enable the machine to organize itself. The problem of how the machine learns is left until we have a clearer idea of what kind of organization needs to be learned.

## 1.2.3. Matrix Models of Associative Memory

By associative memory we mean the ability to get from one internal representation to another or from one part of a complex representation to the remainder. Association has been known to be a prominent feature of human memory since Aristotle. J. R. Anderson and Bower (1973) give a good historic review of the subject in Chapter 3 of their seminal work, *Human Associative Memory*.

Some associations seem to be relatively capricious and unstructured. Others seem regular and predictable. William James (1890-1962) commented that: ''It will be observed that the *object called up may bear any logical relation whatever to the one which suggested it* [p. 284].''

We can easily associate random events, in fact, sophomores have been learning random verbal associations since the dawn of experimental psychology. Yet our awareness that black is the opposite of white or that Robert Kennedy was John Kennedy's brother are examples where two representations are connected by a specific relationship. Selz (1927-1964), in criticizing classical associationism, pointed out that such specific associations are necessary if thought is to be structured and directed rather than degenerating into a diffuse activation of multiple items through nonspecific, pairwise associations. For example, a system cannot retrieve the fact that the opposite of black is white on the basis of

independent pairwise associations between black and white and between opposite and white. Night is just as strongly associated with black and with opposite, so simply combining pairwise effects would activate night just as strongly as white. The specific association needs to be stored as an integrated unit just as white. The specific association needs to be stored as an integrated unit that can be aroused by the combined effects of black and opposite. The question of how integrated units may be stored is treated in some of the subsequent chapters. Here we simply show how pairwise associations could be stored in a parallel system like the brain.

The physiological basis of memory is still something of a mystery. However, there is considerable evidence suggesting that it depends on changes in connectivity between units in a set of interconnected elements. Most neurophysiologists accept that precisely specified changes in synaptic connectivity store memory (see Sec. 1.5). There is clear evidence for this in invertebrates (Kandel, 1976). The suggestion for precise modification that seems most commonly accepted (without detailed physiological evidence, be it noted) is some variant of one originally proposed by D. O. Hebb. Hebb's (1949) suggestion was stated as follows: "When an 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]."

This suggestion predicts that cells will tend to become correlated in their discharges, and a synapse acting like this is sometimes called a correlational synapse. As formulated by Hebb, this model was not suitable for mathematical development. However, in the past few years several groups have developed similar types of parallel, distributed, associative memory which incorporate, in various forms, a learning postulate somewhat like Hebb's. Our feeling is that the qualitative properties shown by existing systems of this kind may be typical of more realistic and complex parallel, associative, distributed systems, so it is worth developing one example of those models in a little detail, including some simple numerical examples.

The models were specifically developed as brain models and as psychological models and have not been used within systems as complex as those studied in artificial intelligence. However, attempts have been made to use the models to generate testable predictions in psychology (Anderson, J. A., 1973).

More detailed explications of parts of these models are available in the original sources and in other chapters in this volume. General references for this section are Kohonen (1977), Willshaw, Buneman, and Longuet-Higgins (1969), J. A. Anderson (1972, 1973, 1977), and Anderson, J. A., Silverstein, Ritz, and Jones (1977). Holographic brain models, which in some respects are similar to the models we describe are developed in Westlake (1970), Willshaw (1971), Cavanagh (1972), and Pribram, Nuwer, and Baron (1974), Willshaw (Chapter 3, this volume) discusses the relationship between holographic and matrix models.

The basic notion in all these models is the idea of a *state vector*, that is, that the currently active representations within the system are coded as patterns of activity simultaneously present on the set of elements that comprise the system. Elements are generally considered to be neurons or very closely related to neurons in these models, and activity is intended to correspond to firing frequency or something very close to firing frequency (e.g., deviation from spontaneous firing rate).

What we wish to store are associations between the state vectors that are the basic entities of the system. But state vectors are not localized in a single place. How can they be handled? Consider a quotation from Karl Lashley (1950) discussed at greater length in Section 1.5. "From the numerical relations involved, I believe that even the reservation of individual synapses for special associative reactions is impossible. [pp. 478–479]."

Lashley argues that there are no privileged sites in the brain for the storage of specific associations in isolation from each other. This idea seems superficially unpromising because it seems that individual associations between pairs of complicated state vectors would interfere with one another if every association used the same set of synapses. To show that interference need not be a problem, we sketch briefly a typical example of such a memory (see Anderson, J. A., 1970, 1972; Cooper, 1974). It is formally a simple linear associator. This model is very similar to those of Willshaw (1971) and Kohonen (1977), which were arrived at independently.

Suppose we have two sets of $n$ neurons, $\alpha$ and $\beta$, which are completely convergent and divergent, that is, every neuron in $\alpha$ projects to every neuron in $\beta$. A neuron $j$ in $\alpha$ is connected to neuron $i$ in $\beta$ by way of a synapse with strength $\alpha$ $(i,j)$. Our first basic assumption, which we have partially justified previously, is that we are primarily interested in the behavior of the set of simultaneous individual neuron activities in a group of neurons. We stress pattern of *individual* activities because our current knowledge of cortical physiology suggests that cells are highly individualistic. For example, their activities at the times of interest to us are typically not correlated with their neighbors, and cell properties differ from cell to cell. We represent these large patterns as state vectors with separate components. We also assume (for this particular model) that these components can be positive or negative. This can occur because the relevant physiological variable in some cases seems to be the deviation in firing rate around a nonzero spontaneous activity level. In other parts of the brain, there may be two separate systems for positive and negative transduction, as in the mammalian visual system with parallel sets of on-center and off-center cells.

Suppose a pattern of activity, **f**, occurs in $\alpha$. Suppose another pattern of activity, **g**, occurs in $\beta$. Suppose that for some reason we wish to associate these two arbitrary patterns. We assume a synaptic modification rule: To associate pattern **f** in $\alpha$ with **g** in $\beta$ we need to change the set of synaptic weights according

to the product of presynaptic activity at a junction with the activity of the postsynaptic cell. Note that this is information locally available at the junction. Thus if $f(j)$ is the activity of cell $j$ in $\alpha$, and $g(i)$ of cell $i$ in $\beta$, then the change in synaptic strength is given by $a(i,j) = \eta\, f(j)g(i)$. We see that this defines an $n \times n$ matrix of changes $\Delta A$ of the form $\Delta A = \eta\, gf^T$ where $f^T$ is the transpose of $f$. Suppose $f$ is normalized, that is, $f.f = 1$, and also $\eta = 1$ so that $\Delta A = gf^T$. Suppose that instead of one association we have $m$ of them, $(f_1, g_1), (f_2, g_2), \ldots, (f_m, g_m)$, each having an incremental matrix $\Delta A_k = g_k f_k^T$. Because there are only the $n^2$ synapses in the system, the same synapses participate in storing all the associations; that is, they are modified again and again. Suppose the overall connectivity is given by

$$A = \sum_k \Delta A_k \qquad (1\text{-}2)$$

By the linearity that, we argue, holds at the synaptic junction, when a pattern of activity, $f_k$, occurs in $\alpha$, it will cause a pattern of activity, $g$, in $\beta$, which is given by $g = Af_k$. Suppose, for the sake of illustration, that patterns $f_k$ are orthogonal. This means that for any pair of patterns $f_l^T \cdot f_k = 0$ for $k \neq l$. If one of the $f_k$ appears at $\alpha$, the activity pattern of $\beta$ is given by

$$g = Af_k = \Delta A_k f_k + \sum_{l \neq k} \Delta A_l f_k$$
$$= g_k + \sum_{l \neq k} g_l(f_l^T f_k)$$
$$= g_k \qquad (1\text{-}3)$$

Thus, for orthogonal $f$'s, the system stores random associations between vectors perfectly.

The capacity of a linear system containing $n$ units and $n^2$ connections is $n$ different associations if the input vectors are orthogonal. If they are nonorthogonal, interference effects become severe as the number of associations approaches $n$. Because the number of neurons in the human brain is of the order of $10^{11}$ this need not be as serious a limitation as it may seem. A single region of the cortex, Area 17, say, may have 50–100 million cells.

Clearly, the activity of a single unit or connection is of little importance to the overall functioning of the system provided the vectors each involve the activity of many different units. On the average, removal of a single unit or connection will cause very slight degradation of many associations rather than complete destruction of a particular one. Wood (1978) has done a number of simulations demonstrating this point.

There is an important property that this simple linear associator has in common with more complicated, nonlinear models that also use a matrix to transform

a state vector. It and its variants are *reconstructive*. This means that the system yields the entire output vector (or a close approximation to it) even if the input is noisy or only partially present, or if there is noise in the memory matrix. This reconstructive property can be used to make a content-addressable memory. First, each of a number of vectors is associated with itself. Then when an incomplete version of one of the stored vectors is used as the input, the output will be a complete or nearly complete version of the vector. So, from any sufficiently large part of the content, the system generates the whole content without using anything like a separate address or entry in an index.

A spectacular example of reconstruction using a linear matrix model is given by Kohonen (Chapter 4, this volume). He shows that the rest of a picture of a face can be reconstructed when the system is presented with a part of the picture. This is an impressive demonstration of the power of the model even if human memory for faces works quite differently. The nonlinear models presented by Willshaw, Hinton, and Anderson and Mozer (Chapters 3, 6, and 8, this volume) show that under certain conditions the output vector can be reconstructed exactly from an incomplete or degraded input.

*A Numerical Example of the Linear Matrix Model.* Table 1.1 shows three input vectors that were chosen to be roughly orthogonal. A required output vector for each of these input vectors was chosen; and the matrix of connection strengths was determined by the three associations between an input and an

TABLE 1.1
The Input Vectors

| Component Number | Input Vectors[a] | | |
| --- | --- | --- | --- |
| | $f_1$ | $f_2$ | $f_3$ |
| 1 | .00 | .00 | -.23 |
| 2 | -.15 | .00 | -.23 |
| 3 | -.29 | -.06 | -.23 |
| 4 | .88 | -.09 | .14 |
| 5 | -.29 | -.30 | .56 |
| 6 | -.15 | .00 | .56 |
| 7 | .00 | .89 | .14 |
| 8 | .00 | .00 | -.23 |
| 9 | .00 | -.30 | -.23 |
| 10 | .00 | -.15 | -.23 |

[a] Correlation between $f_1$ and $f_2$ = .03;
Correlation between $f_2$ and $f_3$ = .06;
Correlation between $f_3$ and $f_1$ = -.02

output vector. The matrix, **A**, is shown in Table 1.2. Formally, it is given by

$$\mathbf{A} = \sum_k \mathbf{g}_k \mathbf{f}_k^T = \mathbf{g}_1\mathbf{f}_1^T + \mathbf{g}_2\mathbf{f}_2^T + \mathbf{g}_3\mathbf{f}_3^T \tag{1-4}$$

Table 1.3 shows three required output vectors, $\mathbf{g}_1$, $\mathbf{g}_2$, and $\mathbf{g}_3$, and also the actual output vectors, $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$, which are slightly different. The difference is caused by the nonorthogonality of the input vectors, which leads to some interference between the different associations.

The effects of local damage or degraded input on this kind of system are illustrated by Kohonen (1977) and Wood (1978) who performed an extensive series of computer simulations of the effects of damage on systems of the type presented here. Wood's results show that, because of chance effects, some elements turn out to be important for particular associations while others are not particularly important to any one association. This corroborates the idea that the individual units in such a system will exhibit varying degrees of specificity.

Table 1.4 shows the effects of degrading the input vector, $\mathbf{f}_3$, by setting some of its components equal to zero.

*A Comparison of Linear and Nonlinear Matrix Models.*   In the linear models the output of a unit is just a weighted sum of its inputs; whereas in the nonlinear models, the output is a more complex function of the weighted sum. The sum may be compared with a threshold to yield a binary value, for example, or it may be rounded up or down if it falls outside certain lower or upper limits. The linear models are easy to analyze and are a sensible and useful first step in the investigation of the whole class of matrix models. However, they have certain insuperable drawbacks that prevent them from exhibiting some of the more interesting properties of the nonlinear models.

If the input and output vectors have the same number of components, it is possible to recycle the output vector and add it to the vector of external inputs to

**TABLE 1.2**
**The Matrix of Weights**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| .055 | .055 | -.033 | -.132 | -.132 | -.033 | .055 | .055 | .055 | .055 |
| .055 | .077 | -.163 | -.088 | -.110 | -.033 | .055 | .055 | .055 | .055 |
| .055 | .098 | .145 | -.288 | -.027 | -.088 | .055 | .055 | .055 | .055 |
| -.033 | -.163 | .098 | .810 | -.156 | -.051 | -.088 | .055 | .064 | .099 |
| -.132 | -.088 | -.288 | -.156 | .491 | .359 | -.185 | -.033 | .055 | .055 |
| -.132 | -.110 | -.027 | -.051 | .359 | .338 | .079 | -.132 | .072 | .143 |
| -.033 | -.033 | -.088 | -.059 | -.185 | .079 | .811 | -.033 | .055 | .055 |
| .055 | .055 | .055 | -.088 | -.132 | -.132 | -.033 | .811 | .055 | .055 |
| .055 | .055 | .072 | -.044 | -.132 | -.132 | -.033 | .055 | .143 | .055 |
| .055 | .064 | -.020 | -.088 | -.132 | -.297 | -.165 | -.132 | .099 | .077 |

form the next input vector to the system. So, given a constant external input vector, the system can run for many iterations. In a nonlinear system it is possible to perform complex computations on the external input vector by repeated iterations using this kind of feedback (see Anderson and Mozer, and Hinton, this volume). In a linear system, however, the result of many iterations through a matrix is just the same as the result of one iteration through some other matrix, so nothing is gained by the multiple iterations.

*Implicit Rules and the Matrix Models.*   In their behavior some matrix models can be shown to act like "rule-governed" systems. There is, however, no explicit representation or application of rules within the system. Matrix models

**TABLE 1.3**
**The Required and Actual Output Vectors**

| Component Number | $g_1$ | $h_1$ | $g_2$ | $h_2$ | $g_3$ | $h_3$ |
|---|---|---|---|---|---|---|
| 1 | 1.00 | 1.00 | .00 | .03 | .00 | -.02 |
| 2 | 1.00 | 1.00 | .00 | .03 | .00 | -.02 |
| 3 | 1.00 | 1.00 | .00 | .03 | .00 | -.02 |
| 4 | 1.00 | 1.03 | .00 | 1.00 | .00 | .04 |
| 5 | .00 | .03 | 1.00 | 1.00 | .00 | .06 |
| 6 | .00 | .03 | 1.00 | 1.00 | .00 | .06 |
| 7 | .00 | .00 | 1.00 | 1.06 | .00 | .06 |
| 8 | .00 | -.02 | .00 | .00 | 1.00 | 1.06 |
| 9 | .00 | -.02 | .00 | .06 | 1.00 | 1.00 |
| 10 | .00 | -.02 | .00 | .06 | 1.00 | 1.00 |

**TABLE 1.4**
**The Required and Actual Responses with Degraded Input**

| Component Number | Required Response to $f_3$ | Actual Response to $f_3$ | Response to $f_3$ with $f_3(6) = 0$ | Response to $f_3$ with $f_3(7) = f_3(8) =$ | Response to $f_3$ with $f_3(9) = 0$ |
|---|---|---|---|---|---|
| 1 | .00 | -.02 | .06 | -.02 | -.02 |
| 2 | .00 | -.02 | .06 | -.02 | -.02 |
| 3 | .00 | -.02 | .06 | -.02 | -.02 |
| 4 | .00 | .04 | .13 | -.15 | -.02 |
| 5 | .00 | .06 | .06 | -.13 | -.15 |
| 6 | .00 | .06 | .06 | -.13 | -.13 |
| 7 | 1.00 | 1.06 | .75 | .68 | -.13 |
| 8 | 1.00 | 1.06 | .68 | .68 | .74 |
| 9 | 1.00 | 1.00 | .68 | .87 | .87 |
| 10 | 1.00 | 1.00 | .68 | .87 | .87 |

have two significantly different types of entity: the active, explicit state vectors, which correspond to activations of the basic units, and the more passive implicit weights in the matrix. It is relatively easy to observe the active, explicit part of the system, but the bulk of the transformational structure is hidden in the implicit part of the system, in the matrices, which are difficult to study because any one weight lumps together many different associations and any one association is distributed over many weights.

It is possible to have complex systems of associations implicit in the matrix of connection strengths. This may cause the system to act in a very rule-governed manner even though there is no process of accessing and applying rules in the sense of a computer program. Because of the interactions between different associations in the implicit structure, adding a set of associations which "agree" in the way they modify a particular subset of the weights may well affect other associations. This transfer of effects to associations that are not explicitly represented may make it appear as if a new rule has been added even though there is no explicit representation of the rule within the system. The status of the rules used in conventional computer models of cognition is thus thrown into question. They may well be *descriptions* of regularities in the behavior of the system that do not correspond, in any simple way, to the representations that are explicit within the system. We return to this issue in Section 1.4.

### 1.2.4.  Minsky's K-Lines Model

Minsky (1980) outlines a theory of the way in which computation may be organized in the human brain. His paper contains a great many speculative suggestions, but the main thrust of his theorizing is that instead of a central processor, which can access arbitrary memory locations by their addresses, the brain may consist of a "society" of fairly simple, local agents, each of which has direct access to a limited number of other agents. Locally, agents may be organized into mutually inhibitory sets, and more globally there will be partial mental states consisting of the currently dominant agents from each local group. Minsky identifies the patterns of active agents with particular mental episodes, and he argues that we need to be able to re-create previous patterns of activity that have proved useful in situations like the current one. To achieve this he proposes that we create a new agent that has connections to all the agents active within a particular mental state. By activating this new agent the old state can be recreated. Minsky elaborates this simple model in various ways to reduce the number of hardware connections required between agents and to allow the re-created state to differ from the old one in ways that make it more appropriate to the situation at hand.

In many respects, Minsky's model constitutes a break with the now-traditional artificial intelligence approach. The agents communicate by emitting excitation and inhibition rather than by passing symbolic expressions, and there is no mention of the problems associated with the creation of new and temporary

object representations at run-time as opposed to just activating old ones. However, the model is an important advance because it takes the brain's hardware seriously. Minsky is particularly concerned with the hardware connections required between agents, a concern partly caused by the technological discovery that, in large-scale integration, it is the connections that cost, not the logical functions (Sutherland & Mead, 1977).

The central idea of the model, that partial mental states are re-created by activating particular agents that designate them, is an interesting intermediate position with respect to the issue of local versus distributed representations. It is the distributed pattern of active agents that is effective in generating the external and internal behavior appropriate to an episode or others like it, but a single local agent can create this pattern of activity. Thus other representations can cause the whole pattern by simply activating that agent. It seems that this combination of local and distributed representations would be effective for representing particular objects or concepts as well as for particular episodes. The advantage of the pattern of active agents as a representation is that new patterns can be created at run-time much more easily than new single agents which require their own specific hardware units and connections.

The real value of Minsky's model, will only be known when the model is specified in sufficient detail for it to be simulated, but the general approach of trying to implement sophisticated computational processes in parallel neuron-like hardware seems extremely promising.

### 1.3.   SYSTEMS OF SIMPLE UNITS WITH FIXED INTERCONNECTIONS

The models reviewed in this section involve systems of interconnected simple units, but they deliberately avoid the issue of how interactions between the units are learned. Instead, the local interactions are specified in advance by the programmer, and the purpose of the model is to demonstrate the computational performance that can be achieved by a system that already has an appropriate set of local interactions between the individual units.

### 1.3.1.  Relaxation Models

Relaxation models typically involve a constraint-satisfaction paradigm in which some input data must be given an interpretation that simultaneously satisfies a large set of local constraints. This interpretation corresponds to a pattern of activity over the units, and it is found by an iterative computation in which each unit affects many others until the whole system settles down into a stable state.

Many of the models are based on a detailed analysis of the computational structure of a specific task. The analysis provides a precise specification of the computational constraints that must be obeyed by any satisfactory interpretation, and these

constraints are then implemented in the local interactions between units. Consequently, the models can only be fully appreciated in the context of the specific tasks that they perform.

Relaxation is best introduced by a classical example. A well-known problem in physics is to calculate the three-dimensional shape of a soap film that is bounded by a nonplanar wire hoop. The shape can be represented by associating a height with each element of a two-dimensional array. The wire hoop fixes the height of the elements at the edge of the soap film. The interior elements obey the constraint that the height of each element is the average of the heights of its neighbors. One way of calculating the heights of the interior elements is to give them arbitrary initial heights, and then to replace simultaneously every interior height by the average of its neighbors. This procedure is called relaxation, and after repeated iterations the heights will settle down to a stable state in which each is the average of its neighbors. This stable state will represent the shape of the soap film.

At least four distinct variations of relaxation have so far been proposed. First, Horn (1977) and Marr (1978) have pointed out that visual systems need to use the intensity information in a raw image to recover the objective characteristics of the surfaces that gave rise to the image. The intensity of each element in the image is the result of many local parameters of the corresponding surface elements. The reflectance of the element, its orientation to the viewer and to the light source, and the level of illumination all interact to determine the image intensity. These interactions can be described by physical equations. In addition, there are normally constraints between the parameters of neighboring surface elements. For example, reflectance usually remains constant, and surface orientation usually changes only slightly from one surface element to the next.

The relationships between local parameters are more complex than the simple neighbor-averaging constraint for the heights of elements in a soap film, but the same kind of relaxation technique can be applied to discover a consistent set of real values for the local parameters of the surface elements causing the image. The intensities in the image act as boundary conditions just like the heights of the elements attached to the wire hoop. The other parameters are given initial values which are successively adjusted to fit the constraints better and better until a stable state is reached. Barrow and Tenenbaum (1978) describe this type of relaxation in more detail.

A somewhat different type of relaxation model was used by Marr and Poggio (1976) to fuse pairs of random dot stereograms (Julesz, 1971). When each eye is presented with one of two random dot patterns, which are identical except for lateral displacement of some regions in one pattern, people see a number of surfaces at different depths. To do this they must decide which dot in one pattern to pair with which dot in the other. Because all dots are the same, there are many potential mates for each one. Each pairing, however, will give a different angular disparity and hence a different perceived depth for the dot. If the assumptions are

made that each dot can only be paired with one other (based on the opacity of surfaces) and that neighboring pairings should have similar disparities and thus be at similar depths (based on the continuity of surfaces), it is possible to make the many potential pairings disambiguate one another.

Marr and Poggio showed that the computation of a good set of pairings can be performed by a machine consisting of multiple units, each of which was a threshold logic unit (see Section 1.2.2). Each unit represented a hypothesis about a particular pairing, and hence it corresponded to a piece of surface at a particular depth. Units corresponding to pieces of surface lying along a line of sight from an eye inhibited one another (the opacity assumption), and units corresponding to adjacent pieces of surface excited one another (the continuity assumption). A dot in a pattern excited all units corresponding to pieces of surface along that line of sight. At each moment a unit is either on or off, and the computation consists of multiple iterations during which units may be turned on or off by the combined effects of the external input and connected units. Whether or not this is a good model of human stereo fusion, it works well for fusing random dot stereograms. It differs from the previous relaxation model in that the values that are adjusted by the relaxation process are binary and correspond to the truth values of hypotheses rather than to continuous properties of surfaces.

A third type of relaxation was introduced by Rosenfeld, Hummel, and Zucker (1976). It is hard to extract information about the contours of objects from an intensity image because some edge segments are locally unclear. Because edges are generally continuous it should be possible to use information from one part of an image to clarify information from unclear neighboring parts. Rosenfeld et al. (1976) suggested that each hypothesis about a local edge should be given an association plausibility between 0 and 1. Initially the plausibilities are determined by how well the particular edge hypothesis fits the local data, but then a relaxation phase is applied during which each hypothesis is affected by its neighbors. The way in which the plausibility of one hypothesis affects the plausibility of another depends on how compatible they are. Rosenfeld and his co-workers have proposed a number of schemes for the interactions (Peleg, 1980), and they have applied their techniques to a number of problems in vision (Rosenfeld, 1978). It is hard to assess just how useful their relaxation technique is because the value of the relaxation phase depends on the extent to which the modified plausibilities are more useful to the higher levels of a vision system than the initial plausibilities. This can only be judged when higher levels exist.

Finally, Hinton (1976, 1977) has proposed a relaxation technique that also associates a value between 0 and 1 with each local hypothesis. Unlike the method of Rosenfeld et al. (1976), however, the hypotheses do not interact directly. Instead, the logical relationships between hypotheses are expressed as numerical constraints, which are implemented as negative feedback loops. Each loop measures the extent to which the constraint is violated by the current values of the hypotheses, and it "tries" to reduce the violation by exerting pressure on the

values of the relevant hypotheses. Each value moves according to the net resultant of the pressures exerted on it by the violated constraints and by the local fit of the hypothesis to the input data. This method has been used to find optimal instantiations of a model in a picture. Unlike the method of Rosenfeld et al. (1976), it is possible to specify precisely what computation is achieved by the relaxation process.

The relaxation techniques are one way of organizing interactions in a parallel machine so that it settles on a good, consistent interpretation of some input data. The drawbacks of relaxation are that it often requires a large number of iterations to achieve equilibrium, and there may be no guarantee that it will find the best solution.

*A Comparison of Relaxation and Nonlinear Matrix Models.* There are many similarities between the formal mechanisms used in relaxation models and those used in the nonlinear matrix models. We have already mentioned the major difference, which is that the relaxation models use weights fixed by the programmer, whereas in the matrix models the weights are determined by the system's experience. Apart from this, however, the main difference seems to reside not in the mechanisms but in the tasks to which they are applied and the interpretation given to the individual units.

Relaxation techniques have typically been applied to low-level vision where it is clear that a great deal of local computation is performed in parallel. The activity levels of individual units are then used to represent the existence of particular local entities or the values of local properties of the surfaces or edges in the visual field. The matrix models, on the other hand, have been applied to problems like recognition and memory, where there is no obvious correspondence between individual units and local properties of the world. Marr and Poggio's (1976) model of stereo fusion, for example, illustrates the close similarity between superficially different parallel models. They use threshold logic units just like perceptrons, and the whole system is equivalent to a nonlinear matrix model but with a sparse matrix.

### 1.3.2. Spreading Activation Models

Collins and Quillian (1972) and Collins and Loftus (1975) argue that the results of certain psychological experiments can be explained in terms of activation spreading along the links of a semantic network in which the nodes represent concepts and the links represent relations between them. They consider tasks like discovering how two concepts are related. The basic idea is that if two different concepts are activated, and activation spreads along the links, then nodes on the path between the two concepts will receive activation from both of them. These nodes will thus have particularly high activation levels.

Other experiments discussed by Collins and Loftus involve semantic priming effects in which activation of one concept tends to speed up reaction times for judgments involving related concepts. In their model similar concepts will have more short pathways between their nodes than dissimilar ones because there will be many other concepts to which both are linked. Thus when a concept is activated, more activation will spread to similar than to dissimilar concepts, and this will explain the priming effect.

The Collins and Quillian model has serious computational deficits. It is hard to use spreading activation effectively for the types of inferential processing for which semantic nets are typically used. There have, however, been two interesting and rather different developments from the Collins and Quillian approach. Fahlman (Chapter 5, this volume) has made it more computationally sophisticated by substantially modifying the idea of spreading activation. He uses discrete markers instead of undifferentiated numerical activation, and this allows him to generate combinations of markers at nodes instead of just activity levels. Fahlman shows that a great deal of computational power and control can be achieved by having a central controller that broadcasts marker-passing instructions to a whole network of nodes, which all obey the instructions in parallel.

In a quite different development McClelland and Rumelhart (1980) have kept the idea of real-valued activation levels, but they have applied it to perception rather than to inference and judgment, and they have been much clearer about the precise rules for propagating activation. Also they have shown how an appropriate scheme for propagating activation can remove the need for a central controller in the particular task domain they have studied. Whether the use of activation for control can be extended to the higher-level tasks that Fahlman tackles remains to be seen.

### 1.3.3. The Rumelhart and McClelland Model of Word Recognition

When a string of letters is presented very briefly, it is easier to recognize the letters if they form a word than if they form a nonsense string (Reicher, 1969). Letter strings which form pronounceable nonwords are intermediate in difficulty. McClelland and Rumelhart (1980) and Rumelhart and McClelland (1980) propose a model in which many simple, neuron-like units interact to produce these effects. For simplicity they restrict themselves to a three-layered system, and they omit feedback from the middle layer to the bottom one.

The bottom layer contains units that detect local features in specific positions within the word. A unit in this layer might, for example, be activated if there is a vertical stroke that could be the right-hand vertical of an H, M, or N in the second-letter position within the word. It is assumed that the feature units occur after some constancy mechanism so that changes in the retinal size, elongation,

position, and orientation of the whole word do not affect the set of feature units that it activates.

In the middle layer a unit represents a specific letter in specific position within the word. Each letter/position unit receives excitatory input from all the feature detectors that fit it and also inhibitory input from feature detectors in the same position that do not fit it.

Units in the top layer represent specific words. Each word unit receives excitatory inputs from all the letter/position units that fit it and inhibitory inputs from the rest. Word units also provide excitatory and inhibitory feedback to the letter/position units. In addition to these interactions between layers there are inhibitory interactions between all pairs of word units and between those pairs of letter/position units that correspond to the same position within the word.

The activity level of a unit is a continuous variable constrained to lie between two limits, and the precise rules for the excitatory and inhibitory interactions and for the thresholds are fairly complex. They are chosen so that when the feature units are activated as they would be by a perceptually presented word, the system settles down into a stable state in which the appropriate word and letter/position units are highly active, and the other units are not.

The impressive achievement of the model is that the precise rules for the interactions can be chosen so that the model is in good agreement with the experimental data for a wide range of experiments. It can, for example, predict the way in which the probability of correctly reporting a particular letter depends on the precise time at which the other letters in the string are presented relative to the letter to be reported.

One intriguing aspect of the model is the way it accounts for the superior recognition of letters in pronounceable nonwords as compared with unpronounceable strings. Units corresponding to pronounceable digrams or trigrams appear to be unnecessary. Letters in a pronounceable string are helped by activity at the word level because there tends to be a whole gang of words which almost fit the string. The combined effects of this gang provide top-down support for each letter, even though every letter is inhibited by the few members of the gang that do not fit it. Although the words within the gang inhibit each other, each pair is in fairly good agreement about the letters, and this agreement causes the system to settle into a stable state in which many word units are slightly active. Thus, pronounceable nonwords are represented by distributed patterns of activity at the word level.

The Rumelhart and McClelland model is rare and promising because it explains experimental data with a computer model that can actually perform the task (given an assumed constancy mechanism). This contrasts with the majority of models in mathematical psychology which merely *describe* the relationships between various aspects of human performance without providing a mechanism for doing the task.

## 1.4.   PARALLEL HARDWARE AND THE SYMBOL PROCESSING PARADIGM

Until recently, most researchers in cognitive psychology and almost all researchers in artificial intelligence have deliberately avoided any serious attempt to specify how their models might be implemented in the brain. There were a number of different reasons for this avoidance. The existing neural network theories were computationally weak—they were incapable of the complex computational tasks that humans routinely perform when they perceive or talk or solve problems. By contrast, computer programs were much more successful at these tasks (e.g., Newell & Simon, 1963; Roberts, 1965; Winograd, 1972).

The major effort in writing programs like these went into software considerations about the representations and processes needed to perform a task. These considerations were largely independent of the particular digital computer on which the software was implemented, mainly because the available digital computers were all extremely similar relative to the range of possible computational machines. Given the implicit assumptions about the computational primitives, the study of the computational properties of complex software was seen, correctly, as a science in its own right, and this distinction between software and hardware was identified with the distinction between the mind and the brain.

Computer programs then transcended their role as a mere tool for implementing theories, and they became metaphors for the mind. Intelligent processes could be implemented in a von Neumann machine, which operated by manipulating abstract symbols according to rules. The symbols themselves did not require any internal structure to give them meaning. Their meaning was determined by the rules for manipulating them, and these in turn were just more symbols. This solved the problem of how thought processes could exist in material objects, and it led to the view, now dominant within cognitive science, that people are symbol-processing machines (Newell, 1980; Pylyshyn, 1980). Conventional digital computers running high-level programming languages are, naturally enough, very good at this kind of processing (much better than people). Hence there is little incentive to investigate radically different kinds of computer architecture.

The symbol-processing approach is supported not just by the fact that computer programs exhibit intelligent behavior, but also by the fact that symbol-processing models provide a remarkably good account of certain human mental processes such as the errors that children make in simple arithmetic (Brown & Burton, 1978) or the verbal protocols that people produce while solving cryptarithmetic puzzles (Newell & Simon, 1972). Any critique of the symbol-processing approach needs to explain why these models work so well.

The models in this volume that use distributed patterns of activity as representations differ from the normal symbol-processing paradigm in an important

way. The internal structure of a symbol is normally thought to be irrelevant to the way it interacts with other symbols. All that is normally necessary is that the symbol have an identity, such as a unique character string, so that it can be compared with others and seen to be either the same or different. The meaning of a symbol is determined by the rules or programs that contain it not by its internal structure.

It is important to realize that this is not the only possible way of organizing a symbol-processing system. It is quite possible for the symbols themselves to have internal structure and for the interactions between symbols to be causally determined by this internal structure rather than governed by stored explicit rules as in the normal symbol-processing paradigm. A symbol, for example, could be a pattern of activity in a large group of hardware units. Provided this pattern is reproducible and regularly causes other such patterns, it is possible to implement symbol processing by the interactions of these patterns.

The symbols in such a system are nonabstract in the following sense: The internal structure of a symbol determines how it interacts with others, so similar symbols tend to have similar interactions. The modifications in the strengths of the hardware connections that are required to alter the causal effects of one symbol will also tend to alter the effects of similar symbols.

Given this view of symbol processing, there are two rather different levels at which a system composed of multiple simple units can be described. At the high level, reproducible patterns of activity can be denoted by abstract symbols, and regular interactions between them can be captured by explicit rules. This kind of a *description* can be implemented rather directly on a conventional digital computer. The impressive performance of programs like those of Brown and Burton (1978) and Newell and Simon (1972) in modeling human behavior are, we think, the result of achieving this level of correspondence between the processes occurring in the brain and the abstract symbol manipulation occurring in the von Neumann machine.

However, because this level omits the internal structure of the symbols and because it captures regularities in causal interactions as explicit rules, there are aspects of cognition for which it is not a good model. For example, the development of the internal structure of the symbols (the specific patterns of activity used to implement them) may not be usefully describable at the same level. Because this internal structure determines how learned effects transfer from one symbol to another, aspects of cognition like the role of similarity and analogy in learning may be outside the appropriate range of the abstract symbol-processing metaphor. We do not mean to imply that these processes cannot be simulated on a von Neumann machine using abstract symbol processing. Indeed, many of the models in this volume are simulated in just this way.

What we are asserting is that the symbol-processing metaphor may be an inappropriate way of thinking about the computational processes that underlie abilities like learning, perception, and motor skills. The rather direct corre-

spondence between the coarse-grained, high level description of our mental processes and abstract symbol manipulation (Card, Moran, & Newell, 1980), may not carry over to the fine-grained description of the highly parallel representations and processes that implement the individual symbols and steps in the coarse-grained description. Just because well-learned and regular interactions between patterns of activity can be captured as explicit rules governing the manipulation of abstract symbols, it does not follow that the emergence of these regularities can be fully captured by models in which explicit rules are added, deleted, or reordered. It is a fallacy to think that the kind of model which works well at one level *must* be applicable at all levels. Rational thought takes years to develop, and it is quite conceivable that it emerges as the highest level of organization of more basic processes that are quite different in character.

To summarize, we are not arguing against the idea that very complex information processing underlies people's ability to perceive, to act, and to learn. Nor are we arguing against computer simulation as a way of exploring such processing. What we are arguing against is the use of abstract symbol manipulation as a prototype for the fine-grained organization of this processing. There are alternative models that have a different computational flavor and that appear to be more appropriate for machines like the brain, which are composed of multiple simple units that compute in parallel.

There are already examples within Artificial Intelligence where hardware considerations have determined general organizational principles. In the early days of computer vision, it was found that it was very difficult to derive a clean line drawing of a scene composed of polyhedral blocks from the mass of gray-level data produced by a camera. Shirai (1973) showed how higher-level knowl-edge could be used to guide line finding so that the computer could restrict its slowest and most accurate line-finding techniques to areas of the image likely to contain lines. Shirai's program was used to support the idea that really competent vision systems require rich interactions between experts in different domains (like line finding and shape representation) rather than being restricted to a pass-oriented organization, in which each level of processing is uninfluenced by subsequent levels.

The application of this idea to the early stages of visual information processing was attacked by Marr (1976) who argued that the highly parallel hardware known to exist in the brain could produce much richer representations of edges and local surface elements than existing Artificial Intelligence programs without invoking knowledge of particular objects. The dispute has not been fully settled, but there seems no doubt that much of the plausibility of Marr's theory stems from the existence of a great deal of parallel hardware in the brain that is devoted to early visual processing.

The availability of parallel hardware drastically changes arguments about the relative efficiencies of different computational algorithms. The total number of computational operations becomes less important than the question of whether

the operations can be performed in parallel or whether there is a necessarily sequential structure in which one operation cannot be performed until the results of others are known.

## 1.5. PARALLELISM AND DISTRIBUTION IN THE MAMMALIAN NERVOUS SYSTEM

This section outlines the anatomy and physiology of the mammalian neocortex. More details can be obtained from books such as Shepherd (1979), Brazier and Petsche (1978), and Kuffler and Nicholls (1976). Some striking themes are apparent even at the superficial level that we can discuss the system. There is clear evidence, for example, of parallelism and a degree of distribution.

Most of the visible bulk of the human brain is neocortex. Grossly, the cortex is a two-dimensional folded sheet, consisting of the gray matter, a 3–6 mm layer on the outside of the brain containing the cell bodies, and the white matter, a tremendous number of incoming and outgoing fibers. The surface area of the cortex is around 1 square meter, but the sheet is so folded and convoluted— presumably for compact packing—that only about a third of the cortex is visible from the outside; the rest is submerged in fissures.

There are two seemingly contradictory main themes that characterize cortical organization: differentiation and homogeneity. Although there are clear differences between different areas of the cortex, the basic cell types and the fundamental organization of all parts of the neocortex are surprisingly similar. The basic circuitry seems to be the same everywhere.

There are generally held to be two broad classes of neurons in the cerebral cortex: pyramidal cells and stellate cells, with the stellate cells containing a number of different subgroups. The two-dimensional sheet of gray matter of the cortex is itself strongly layered. Neuroanatomists generally identify six layers. The pyramidal cells send their axons to other parts of the cortex and to other regions of the central nervous system. They thus form the "output" cells of a region of cortex through there is no clear flow of information from input to output in the neocortex, a property that complicates analysis. The stellate cells are generally smaller and send axons to a circumscribed local region of the cortex, presumably being primarily short-range "intrinsic" cells.

Pyramidal cells can be very large, with a typical pyramid-shaped cell body and with a large "apical" dendrite, which runs to the surface of the cortex, through the layers, and perpendicular to the cortical surface (see Fig. 1.1). The cortex is a remarkably Cartesian system, with layers parallel to the surface and with fibers and dendrites running perpendicular to the layers and cutting across them. As has been pointed out on numerous occasions, this is in no sense a random network but is exquisitely structured, both anatomically and, as we are now discovering, physiologically.
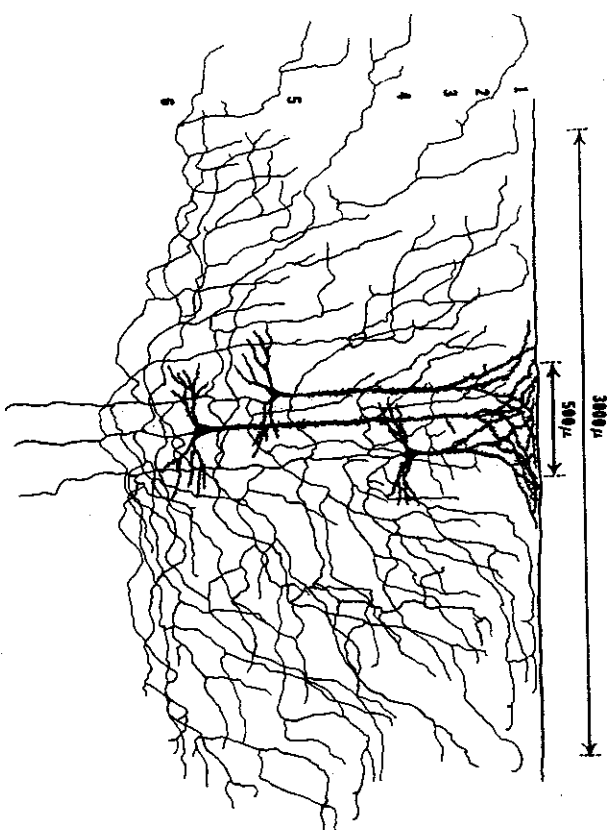
FIG. 1.1.   Three pyramidal cells in cerebral cortex, surrounded by recurrent collaterals. The thin, fuzzy processes on the pyramidal cell dendrites are dendritic spines. The outer surface of cortex is at the top; the apical dendrites run perpendicular to the surface. Drawn from several sections of 60-day cat cortex stained by a rapid Golgi variant. (Scheibel & Scheibel, 1970).

The apical dendrites run to the surface of the cortex and then branch. Other dendrites come off the base of the pyramid forming a number of "basal" dendrites. The extent of branching of these dendrites forms a cylinder which seems to be an important information-processing module of the cortex. The apical dendrite of large pyramidal cells whose cell body is in layer 5 may be several millimeters in length. This cell may receive extensive synaptic input over its entire length. Traditional estimates of numbers of synapses on cortical pyramids run from about 7000 in the visual cortex to 50,000 in the motor cortex.

A notable feature of pyramidal cells is the presence of small processes, a few microns in length, called "spines." At one time they were thought to be artifacts of the neuroanatomical staining methods used to visualize the cells, but now they are known to be real and important. All synaptic contacts outside the cell body are made on spines.

It has not been possible to show synaptic modification in vertebrates in detail or to have any idea of the quantitative form of the modification, if present, but the data and theory suggest that spine modification is one likely candidate for the detailed synaptic changes that almost certainly underlie learning in adults. The

structure of the spine is ideally suited to the kind of correlational modification scheme proposed by Hebb (1949), in which synaptic changes depend on the conjunction of presynaptic and postsynaptic activity. The presynaptic element is at one end of the spine. A dendrite much thicker than the spine forms a low-resistance pathway to the cell body. Thus in close physical proximity, separated by only a few microns, we have presynaptic and postsynaptic activity. Rall and Rinzel (1973) have shown that the anatomy of the spines allows easy modification of synaptic coupling: Slight changes in length or thickness can cause substantial changes in degree of coupling between presynaptic and postsynaptic elements, and normal spines fall into the range where such sensitivity to modification is predicted theoretically.

In immature organisms considerable evidence suggests that change in the amount of dendritic branching is important as well, and dendrites and spines have been shown to respond to environmental influences in both the cortex and the cerebellum (Floeter & Greenough, 1979; Globus, Rosenzweig, Bennett, & Diamond, 1973; Pysh & Weiss, 1979; Volkmar & Greenough, 1972).

The axons of pyramidal cells branch extensively and reenter the cortex up to 3 mm away from the cell of origin. Szentágothai (1978) has suggested that such recurrent collateral connections are "quasi-random," synapsing with the first suitable candidate to be in their way. Such collateral connections generally contact other pyramidal cells and the contacts seem to be excitatory though this has been difficult to show physiologically (see Fig. 1.2).

At the level of the afferent connections of neurons, Szentágothai (1978) comments:

There cannot be much doubt today that afferent input to cortex . . . is geometrically highly ordered. . . . The basic principle of order appears in many cases to be simply some parallel lamination of fibers of common origin or some other feature that they have in common. In other cases the ordering principle is more complex and may be the preservation of an almost complete isomorphism in the cortical representation of the periphery [p. 77-78].

The picture emerging from such considerations is one of a very high degree of specific wiring both in distant and in local connections of the cerebral cortex . . . [pp. 81].

The one exception to this rule may be the collaterals which may exhibit a degree of 'randomization' that would give room for fortuitous connections. We can conclude that the cortex is a highly ordered structure with the possibility of some salutary local chaos.

A striking observation about cortical circuitry is that it is possible to speak about "cortical circuitry." The basic cell types and circuits seem, with relatively minor variations, to be similar everywhere in the neocortex. Variations exist in the thickness of the cortex, the number and arrangement of layer 4 stellate cells,
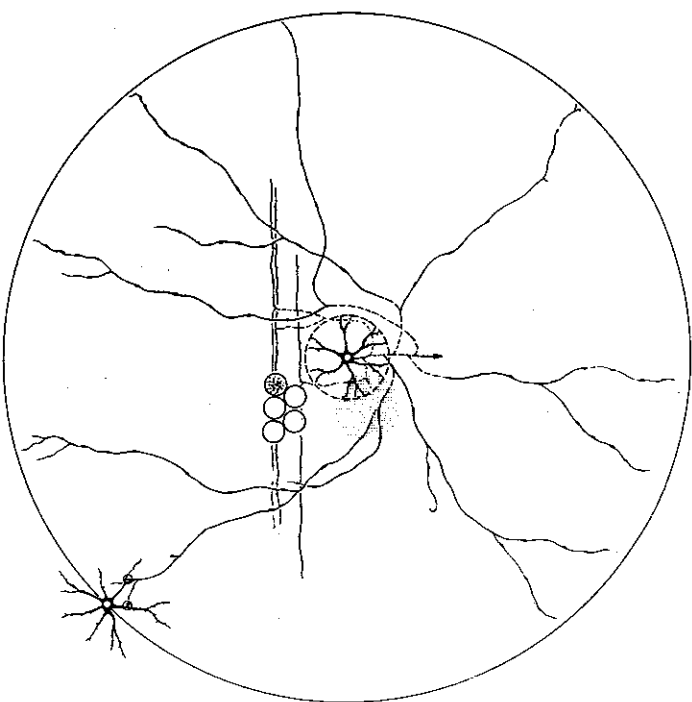
FIG. 1.2.   Collateral system of a pyramid cell (center) as it would appear in the view from the surface (semidiagramatic reconstruction). The large circle having a diameter of 3 mm indicates the potential territory reached by the collaterals. The pyramid cell below right on the perimeter would be a potential recipient of synaptic contacts from the central cell. The small circles (100-μ diameter) correspond to the apical dendrite bundles containing 20-30 apical dendrites of a pyramid cell cluster. The round stippled area at right of central pyramid cell (±300-μ diameter) indicates width of an "arborization-column" of corticocortical afferents. (Szentágothai, 1978).

for example, or the occasional appearance of large and striking variants of a cell type, such as the giant Betz cells (pyramidal cell variants) in the motor cortex. However the basic plan, the arrangements in columns, in layers, and the same typical connections seem to be everywhere (Shepherd, 1979).

On the basis of rather small differences in structure Brodmann distinguished about 50 different cortical areas; other anatomists have made slightly different distinctions. Such distinctions are not mere parcellation. They often describe functional specialization. For example, Area 17 of the primate cortex is distinguished by a prominent fine white line called the "stria of Gennari." Area 17 is also called "primary visual cortex" (sometimes striate cortex) because the lateral geniculate body sends its most prominent projection to Area 17 in mammals. We

should emphasize that these subareas of cortex and related subcortical areas are connected together in complex ways, in series, in parallel, and with potential loops. Figure 1.3 shows a partial picture of the earlier stages of visual system connectivity. The details of the projections of the visual system are of considerable interest because they illustrate an especially striking example of the precision of connection of the cortex.

Some levels of the visual system are intrinsically a two-dimensional parallel system: A layer of receptors in the retina projects, after intensive local processing involving important lateral effects, to a parallel array of retinal ganglion cells. The axons of the million or so ganglion cells go up the optic nerve to the lateral geniculate. The lateral geniculate is a six-layered structure in primates, with the projection from the ipsilateral eye (same side) occupying three layers, and the projection from the contralateral (opposite side) eye occupying the other three. The projection is not random but very precise, so that corresponding points in the visual fields of the two eyes, though still separate, are brought into register above each other in different layers, and a spatial map is maintained. A distorted (but not torn or dislocated) map of visual space is present in the geniculate. This projection maintains local continuity and topography. The projection from geniculate to cortex gives rise to a system that also has a precise map. The work
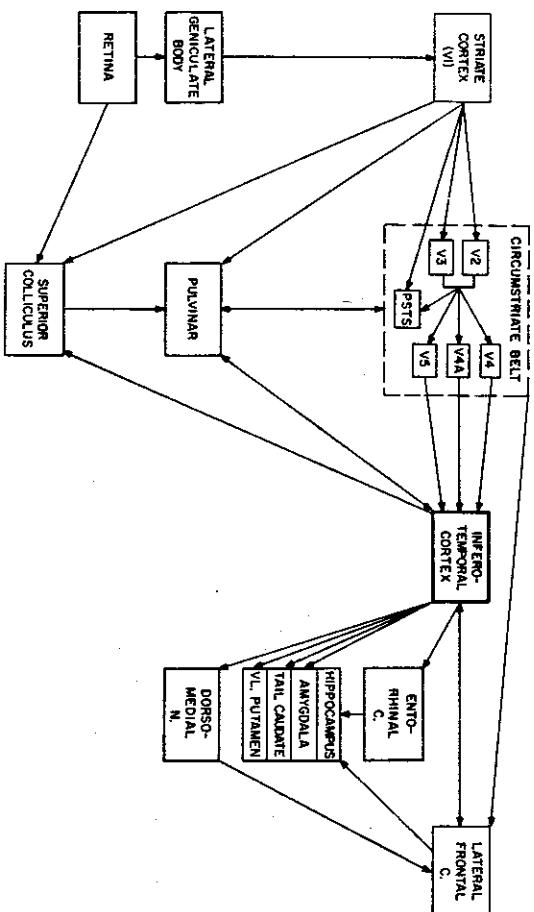


FIG. 1.3. Major ipsilateral afferent connections of inferotemporal cortex in macaque monkey, *Macaca mulatta*. This figure shows the complexity of pathways involved in a portion of the visual system connecting to the highest "visual" cortical region, inferotemporal cortex, which seems to be involved in complex visual function. (Gross et al., 1974).

FIG. 1.4. Reconstruction of the ocular dominance pattern over the entire exposed part of the right primary visual cortex, from a series of sections stained by a reduced silver method developed by Simon LeVay. (LeVay, Hubel, & Wiesel, 1975). The left hand margin is at the medial edge of occipital lobe, where cortex folds downward; the area of cortex shown is roughly 2 cm in extent. (Hubel & Wiesel, 1979).

of Hubel and Wiesel and collaborators has shown that the different eyes project to what are essentially bands of eye dominance in cortical layer 4 which are brought together in the other layers so that cells outside of layer 4 are typically binocular. This is strikingly demonstrated in Fig. 1.4, which shows the alternating bands of eye dominance in cat visual cortex (Area 17).

As is well known, cells in Area 17 in the cat and monkey respond preferentially to oriented line segments. All the cells in a single cortical column seem to have the same orientation. The preferred direction of orientation shifts at a constant rate across the surface of the cortex. Bands of constant orientation intersect the bands of ocular dominance. Although there is precise mapping in that, at the scale of millimeters, there is a good average spatial map on the cortex, there is considerable noise and jitter in a small region, so a single cortical column may contain cells that respond to slightly different parts of the visual field (Albus, 1975). The diversity of single-cell response and the observation that nearby cells in the neocortex may be quite unlike each other in details of behavior, though corresponding in general properties such as (in the visual system) location in the visual field or orientation, seems well established (see Fig. 1.5). A
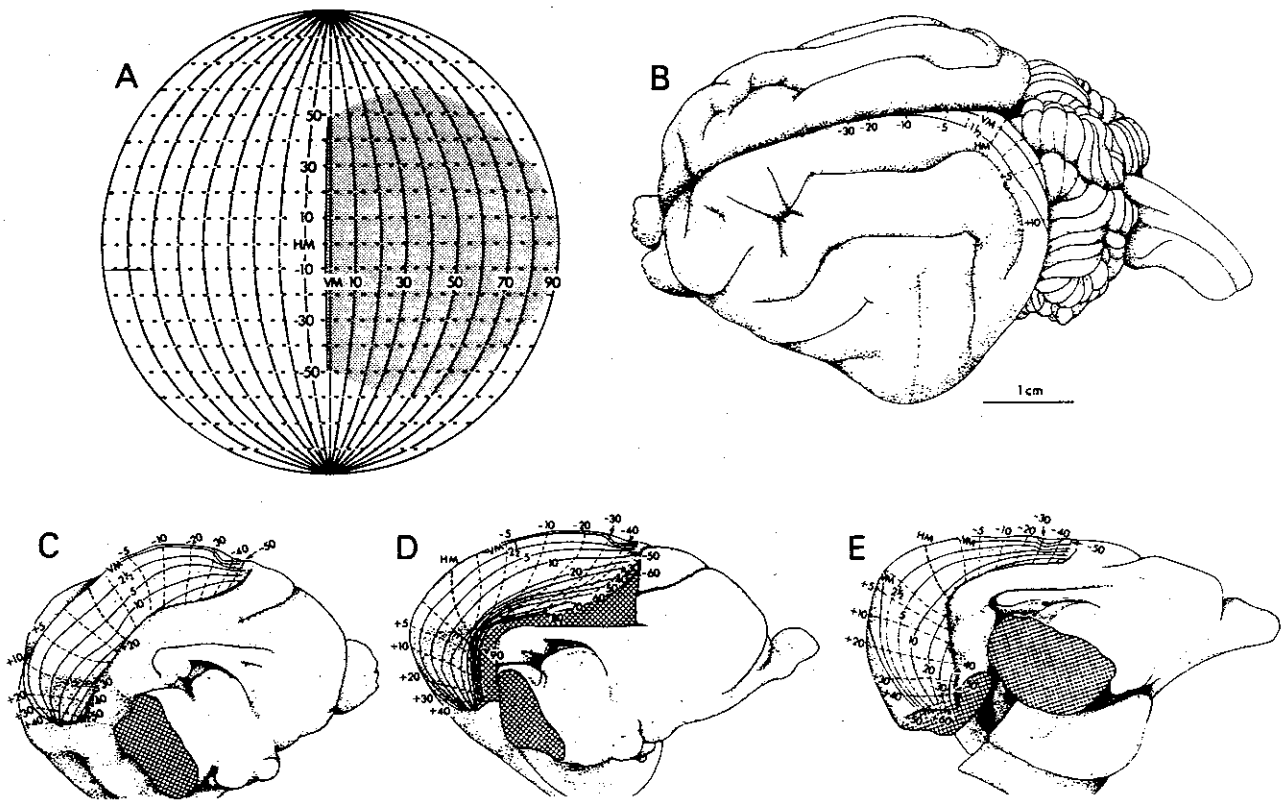
FIG. 1.5. Diagram of the representation of the visual field in Area 17.A is a perimeter chart showing the extent of the visual field represented in Area 17. The chart is based on a world co-ordinate scheme in which the azimuths are represented as solid lines and the elevations as dashed lines. The location of the visual field in Area 17 is illustrated in the four sketches of cat brain shown in B–E. (Tusa, Palmer, & Rosenquist, 1978).

fine discussion of the visual cortex with truly striking pictures is easily available (Hubel & Wiesel, 1979).

The precise topographic maps that are typical of visual inputs seem to be ubiquitous. At present we have many maps representing the sensory receptors on the body surface. For the motor cortex there is the famous "motor homonculus," a drawing of which is a staple of introductory psychology books, and for the auditory system there is a map in terms of frequency. These maps seem in general to conform to the rule of precision on the average but jitter in detailed local characteristics. Multiple and overlapping maps also seem common. At present multiple visual maps have been located in cats, all apparently analyzing the visual input in different ways (Zeki, 1978). An example has also been found of a case in the circumstriate visual cortex where a map of visual space and a map of auditory space are in register, so a cell might be excited by a visual or by an auditory stimulus if it is in the appropriate spatial location (Morrell, 1972).

Another aspect of these maps that is theoretically important is their distortions. The primate retina contains an area of relatively high optical quality that is very rich in receptors, the fovea. This area is highly overrepresented in number of receptors, retinal ganglion cells, and consequently nerve fibers associated with it. This preeminence is maintained all the way to the primary visual cortex, where over half of the cortical neurons are concerned with analyzing only the few degrees of visual space represented by the fovea and the area around it, and the remaining portion of the visual cortex is concerned with the remainder of visual space.

It seems to be a general rule that the more important a sensory system for the animal's behavior, the larger its relative cortical representation. A well-known example is the human motor homonculus, which has disproportionately large hands compared to the feet because hands are much more important in our behavior than feet. The rhesus monkey "monkeyunculus" has roughly equal-sized hand and foot representations. This "mass effect" is confirmed over and over in cortical organization. A particularly striking example is the bat auditory cortex. Often the distortions undergone by maps are apparently quite lawful; for example, a roughly logarithmic transformation seems to be found in both the visual system of the cat (Fischer, 1973) and in the tonotopic organization in the auditory cortex. However, telling exceptions occur. The mustache bat is an echo locator, and the maps of auditory space on its auditory neocortex seems to show clearly the use of map distortions and constructions as important aspects of information processing. The mustache bat emits cries with strong frequency components at 61.0 and 91.5 kHz. There are "disproportionately large" portions of this bat's cortex devoted solely to these two frequencies and a relatively small representation of other frequencies, causing a great distortion of the tonotopic map (Suga & Jen, 1976, as shown in Fig. 1.6).

The mustache bat uses a fairly complicated call (see inset on Fig. 1.7C). The call contains a previously mentioned constant frequency portion that allows a
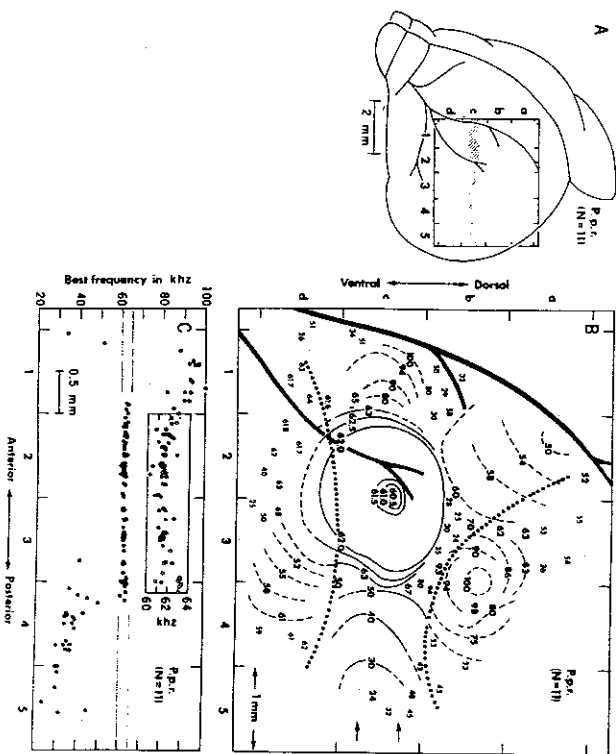
FIG. 1.6. (A) Dorsolateral view of the mustache bat cerebrum. The auditory cortices are within the rectangle. (B) Distribution of best frequencies in the rectangle shown in (A). The area between the dotted lines is the primary auditory cortex (AI). The areas dorsal or ventral to the AI are nonprimary auditory cortices. Orderly tonotopic representation is clear in the areas with dashed contour lines, but it is vague in the areas with dashed contour lines. In the areas where contour lines are not drawn, the tonotopic representation, if present, is obscure. Some of the best frequencies obtained in the obscure areas are shown by small-print nos. (C) Distribution of the best frequencies along the anteroposterior axis in the shaded area in (A). Since the minor differences among the best frequencies in areas 2 and 3 cannot clearly be shown in (C), the distribution of best frequencies in this area is shown by the inset with a larger frequency scale and using open circles. P.p.r.: Pteronotus parnellii rubiginosus (species name). (Suga & Jen, 1976).

return echo to be analysed for Doppler shift to obtain relative velocity information. There is also a brief period at the end of the call where the emitted frequency suddenly drops several kilohertz in a few msec. This portion of the call, referred to as "chirp," by radar engineers, allows range information to be computed because the time at which a received signal was emitted can be computed from its frequency. Chirping is also an optimal technique for several other reasons having to do with the energy requirements of the emitted signal. However the computations required to obtain range information are not trivial, and in radar it can be computed with a dispersive delay line. It is remarkable that this bat has and co-workers (Suga & O'Neill, 1979) have demonstrated that this bat has

constructed, on the surface of its cortex, a map for target range, a derived quantity not immediately present in the signal.

The apparent importance, necessity, and universality of maintaining and even constructing a two-dimensional quasi-continuous representation of important aspects of the environment is a significant clue to the kind of parallel processing being performed by the cortex. We must emphasize, however, that maps continuous on the average and overall are not continuous in very small areas. Slabs of eye dominance intersected by orientation columns give rise to discontinuous
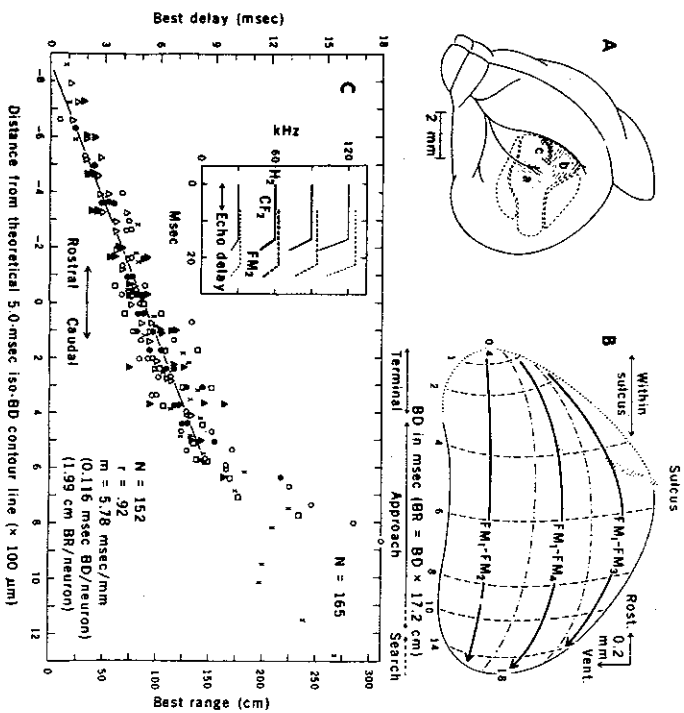


FIG. 1.7. (A) The left cerebral hemisphere of the mustache bat, showing (a) Doppler-shifted CF (= constant frequency), (b) FM (= frequency modulation), and (c) CF/CF processing areas. (B) The FM processing area consists of three major clusters of delay sensitive neurons: FM$_1$–FM$_2$, FM$_1$–FM$_3$, and FM$_1$–FM$_4$ facilitation neurons. Each cluster shows odotopic representation (i.e., topographical representation of target range). Range information in the search, approach, and terminal phases of echolocation is represented by activity at different loci in the cerebral hemisphere. (C) The relation between BD (best delay) and distance along the cortical surface. The data were obtained from six cerebral hemispheres and are indicated by six different symbols. The inset is a schematized sonagram of an orientation sound and a Doppler-shifted echo in the approach phase of echolocation. (Suga & O'Neill, 1979).

local representations of a continuous physical stimulus, a point emphasized by Hubel and Wiesel (1979).

Let us sum up the aspects of cortical neuroanatomy that are of importance to us here because they seem to be generalizations that may lead to approximations of theoretical interest:

1. The cortex is precise on the average in its connections both in terms of its incoming afferent projections (maps) and in terms of its internal connections. Topographic aspects of the external world may be maintained or reconstructed with considerable accuracy.

2. The cortex displays a very strong mass effect so that areas important in an animal's behavior are exaggerated proportionately.

3. The cortex displays precision on the average but imprecision in small areas, in that single neurons may deviate considerably from the average. A certain random component (i.e., Szentágothai's comments about the recurrent collaterals) may be present along with a more precise component. Neurons display considerable individuality, though partaking of some average similarities to neighbors.

4. The cortex is parallel in its organization from the initial afferent inputs, to its layered structure, to its topographically organized maps. Within a given region, such as Area 17, say, parallel organization is very striking. Serial connection of cortical regions, one region to the next, is present, but the individual cortical areas are highly parallel in their organization.

The evidence for parallelism of large aspects of cortical function is overwhelming. The evidence for distribution is also strong but more controversial. It is an experimentally testable area with extreme views on both sides present in the literature.

*Neuron Specificity.* There are two extreme positions that one can take. One can claim great specificity and importance for single neurons. We can look at the physiological data and observe that single neurons respond to only a small number of stimuli; that is, they have considerable selectivity. We could then conjecture that when a neuron is active it signals very precise information about the sensory input. This point of view is held by Horace Barlow (1972), who summed up his position in a series of dogmas, three of which are particularly relevant here:

Dogma 2 states: "The sensory system is organized to achieve as complete a representation of the sensory stimulus as possible with the minimum number of active neurons [p. 371]."

Dogma 4 states: "Perception corresponds to the activity of a small selection from the very numerous high-level neurons, each of which corresponds to a

pattern of external events of the order of complexity of the events symbolized by a word [p. 371]."

Dogma 5 states: "High impulse frequency in such neurons corresponds to high certainty that the trigger feature is present [p. 371]."

Thus: Single neurons are very important, very specific and signal certainty with increased activity.

The real virtue of this position is that it is easy to understand and makes good intuitive sense. When a cell fires, something specific and important happens, both because very few other cells are talking and because the cell is "meaningful." This position is held, deep in their hearts, by many practicing neurophysiologists because it seems to give a clear interpretation of what they observe with their microelectrodes.

*Distribution.* Karl Lashley, on the other hand, is identified with a strong statement of distribution. Although most of his work was concerned with the problem of memory, his ideas on distribution are quite general. To quote from the summary of his famous 1950 paper, "In Search of the Engram":

It is not possible to demonstrate the isolated localization of a memory trace anywhere within the nervous system. Limited regions may be essential for learning or retention of a particular activity, but within such regions the parts are functionally equivalent. The engram is represented throughout the area.... Briefly, the characteristics of the nervous network are such that when it is subject to any pattern of excitation, it may develop a pattern of activity, reduplicated throughout an entire functional area, by spread of excitations, much as the surface of a liquid develops an interference pattern of spreading waves when it is disturbed at several points..... Consideration of the numerical relations of sensory and other cells in the brain makes it certain, I believe, that all of the cells of the brain must be in almost constant activity, either firing or actively inhibited. There is no great excess of cells which can be reserved as the seat of special memories. The complexity of the functions involved in reproductive memory implies that every instance of recall requires the activity of literally millions of neurons. The same neurons which retain the memory traces of one experience must also participate in countless other activities.

Recall involves the synergic action or some sort of resonance among a very large number of neurons.... From the numerical relations of individual synapses for special associative reactions is impossible [pp. 477-480].

With this diversity of views, one can now look at the experimental literature and decide how the observed neuronal specificities agree with these two positions. Clearly, there is specificity. Cells do not respond to all conceivable stimuli or even a small subset of them but are quite specific in their responses. At the

same time they are not, in our opinion, so specific as to be what Barlow's dogmas would lead one to expect. In fact, we conjecture if this were the case that single unit neurophysiology would be hopeless once beyond the very lowest levels of cortex because it would be almost impossible to ever find the precise stimulus required to drive any given cell. What we seem to observe in actuality is a spectrum of response types from quite specific and narrowly tuned to quite general and broadly tuned. As one well-known example, in inferotemporal cortex, a higher level visual center, many cells are quite broadly tuned, responding to orientation over many degrees of visual angle. Yet some cells (e.g., the famous "monkey hand" cell) seem to be very specific. Yet even the "monkey hand" cell can be driven by other stimuli such as some geometric shapes, while reserving its highest frequency discharge for the shape corresponding to a silhouette of a monkey hand (Gross, Bender, & Rocha-Miranda, 1974).

We suggest that truth lies somewhere in the middle of the two extreme views and that there is a moderate amount of distribution in the cortex, so that any single cell responds to many things but nowhere near all things. It seems to us both experimentally observed and sensible from the point of view of information processing that there be a considerable range of specificities from quite specific to quite broad.

Singer (1978), in an analysis of the visual cortex, comments that:

The data . . . suggests that the result of any higher level integrative operation is presumably not encoded in the specific responses of a few but highly selective cells. . . . This might indicate a cooperative principle of encoding whereby the message about the presence of a particular combination of features is conveyed by the graded, mutually dependent responses of a large number of cells within a functional matrix of cells [p. 377].

One of us has discussed this question of single-unit specificity elsewhere (Anderson, J. A., et al., 1977), and Singer's remarks seem to us to capture the actual situation.

## 1.6.  SUMMARY

We have presented a necessarily limited selection from among the wide range of models in which computation is performed in parallel by multiple, rather simple units. We feel that such models are worth further study for several quite different reasons.

First, the neocortex is clearly highly parallel in the arrangement of its basic elements. A single cortical area contains many millions of neurons arranged and connected with a high degree of parallelism. It seems entirely reasonable that such pronounced structure should have implications for cognitive function.

Second, there are a number of functional attributes of cognition, ranging from memory access and concept formation to stereopsis and visual perception, that seem to be very hard to explain with more traditional serial models. The existing models using parallel systems of simple units have already had some success in these areas and more can be expected as we gain further insight into the properties of this kind of computation. We feel that distributed representations are a very promising idea and that a successful cognitive theory, when it comes, may take the form of a calculus of state vectors where the psychological level of description will correspond to the permutations and interactions of the state vectors.

Finally, there are purely technological reasons for studying parallel systems, outside of their scientific interest, because they may be the best way to increase the speed and power of computation in the future. Recent developments in Very Large Scale Integration (VLSI) and in Computer Aided Design (CAD) make it relatively easy to implement parallel models directly in the hardware, expecially if the models use simple units and regular interconnection schemes.

Many of the systems we have discussed are still somewhat unfamiliar and difficult to work with. Parallel models tend to have intrinsic characteristics of their own, and the nonlinear ones are often very hard to analyze. Considerable experience with them may be needed before our ideas about parallel computation are adequate to allow us to exploit the enormous potential of this class of systems both as computational devices and as models of the mind.

## REFERENCES

Albus, K. A quantitative study of the projection area of the central and paracentral visual field in Area 17 of the cat. 1. The precision of the topography. Experimental Brain Research, 1975, 24, 159-179.

Anderson, J. A. Two models for memory organization using interacting traces. Mathematical Biosciences, 1970, 8, 137-160.

Anderson, J. A. A simple neural network generating an interactive memory. Mathematical Biosciences, 1972, 14, 197-220.

Anderson, J. A. A theory for the recognition of items from short memorized lists. Psychological Review, 1973, 80, 417-438.

Anderson, J. A. Neural models with cognitive implications. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review,* 1977, *84,* 413-451.

Anderson, J. R. & Bower, G. H. *Human associative memory.* Washington, D.C.: V. H. Winston, 1973.

Barlow, H. Single units and sensation: A neuron doctrine for perceptual psychology. *Perception,* 1972, *1,* 371-394.

Barrow, H. G., & Tenenbaum, J. M. Recovering intrinsic scene characteristics from images. In A. R. Hanson & E. M. Riseman (Eds.), *Computer vision systems.* New York: Academic Press, 1978.

Brazier, M. A. B. & Petsche, H. (Eds.) *Architectonics of the cerebral cortex.* New York: Raven, 1978.

Brown, J. S. & Burton, R. R. Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science,* 1978, *2,* 155-192.

Card, S. K., Moran, T. P., & Newell, A. Computer text-editing: An information-processing analysis of a routine cognitive skill. *Cognitive Psychology,* 1980, *12,* 32-74.

Cavanagh, P. Holographic processes realizable in the neural realm: Prediction of short-term memory performance (Doctoral Dissertation, Carnegie-Mellon University, 1972.) *Dissertation Abstracts International,* 1973, *33,* 3280B.

Collins, A. M., & Loftus, E. F. A spreading activation theory of semantic processing. *Psychological Review,* 1975, *82,* 407-428.

Collins, A., & Quillian, M. R. Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley, 1972.

Cooper, L. N. A possible organization of animal memory and learning. In B. Lundquist & S. Lundquist (Eds.), *Proceedings of the Nobel symposium on collective properties of physical systems.* New York: Academic Press, 1974.

Fischer, B. Overlap of receptive field centers and representation of the visual field in the cat's optic tract. *Vision Research,* 1973, *13,* 2113-2120.

Floeter, M. K. & Greenough, W. T. Cerebellar plasticity: Modification of Purkinje cell structure by differential rearing in monkeys. *Science,* 1979, *206,* 227-229.

Globus, A., Rosenzweig, M. R., Bennett, E. L., & Diamond, M. C. Effects of differential experience on dendritic spine counts in rat cerebral cortex. *Journal of Comparative and Physiological Psychology,* 1973, *82,* 175-181.

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. Inferotemporal cortex: A single unit analysis. In F. O. Schmitt & F. G. Worden (Eds.), *The neurosciences: Third study program.* Cambridge, Mass.: MIT Press, 1974.

Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.

Hinton, G. E. Using relaxation to find a puppet. *Proceedings of the A.I.S.B. Summer Conference,* University of Edinburgh, 1976.

Hinton, G. E. *Relaxation and its role in vision.* Unpublished doctoral dissertation University of Edinburgh, 1977.

Horn, B. K. P. Understanding image intensities. *Artificial Intelligence,* 1977, *8,* 201-231.

Hubel, D. H. & Wiesel, T. N. Brain mechanisms of vision. *Scientific American,* 1979, *241*(3), 150-162.

James, W. *Psychology: Briefer course.* New York: Collier, 1962. (Originally published, 1890).

Julesz, B. *Foundations of cyclopean perception.* Chicago: The University of Chicago Press, 1971.

Kandel, E. R. *Cellular basis of behavior.* San Francisco: W. H. Freeman, 1976.

Kohonen, T. *Associative memory: A system-theoretical approach.* Berlin: Springer, 1977.

Kuffler, S. W. & Nicholls, J. G. *From neuron to brain.* Sunderland, Mass.: Sinauer Associates, 1976.

Lashley, K. In search of the engram. In Symposia of the Society for Experimental Biology, No. 4., *Physiological mechanisms in animal behavior.* New York: Academic Press, 1950.

LeVay, S., Hubel, D. H., & Wiesel, T. N. The pattern of ocular dominance columns in macaque visual cortex revealed by a reduced silver stain. *Journal of Comparative Neurology,* 1975, *159,* 559-575.

McClelland, J. L. & Rumelhart, D. E. *An interactive activation model of the effect of context in perception: Part I.* (Tech. Rep. 91). San Diego: University of California, San Diego, Center for Human Information Processing, 1980.

McCulloch, W. S. & Pitts, W. H. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics,* 1943, *5,* 115-133.

Marr, D. Early processing of visual information. *Philosophical Transactions of the Royal Society, Series B,* 1976, *275,* 483-524.

Marr, D. Representing visual information. In A. R. Hanson & E. M. Riseman (Eds.), *Computer vision systems.* New York: Academic Press, 1978.

Marr, D. & Poggio, T. Cooperative computation of stereo disparity. *Science,* 1976, *194,* 283-287.

Minsky, M. Plain talk about neurodevelopmental epistemology. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence.* Boston, Mass. 1977.

Minsky, M. K-lines: A theory of memory. *Cognitive Science,* 1980, *4,* 117-133.

Minsky, M. & Papert, S. *Perceptrons.* Cambridge, Mass.: MIT Press, 1969.

Morrell, F. Integrative properties of parastriate neurons. In A. G. Karczmar & J. C. Eccles (Eds.), *Brain and human behavior.* New York: Springer, 1972.

Neimark, E. D. & Estes, W. K. (Eds.). *Stimulus sampling theory.* San Francisco: Holden-Day, 1967.

Newell, A. Physical symbol system. *Cognitive Science,* 1980, *4,* 135-183.

Newell, A., & Simon, H. A. GPS, a program that simulates human thought. In E. Feigenbaum and J. Feldman (Eds.), *Computers and thought.* New York: McGraw-Hill, 1963.

Newell, A., & Simon, H. A. *Human problem solving.* Englewood Cliffs, N.J. Prentice-Hall, 1972.

Nilsson, N. J. *Learning machines.* New York: McGraw-Hill, 1965.

Norman, D. A., & Bobrow D. G. Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology,* 1979, *11,* 107-123.

Peleg, S. A new probabilistic relaxation scheme. *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence,* 1980, *362-369.*

Pribram, K. H., Nuwer, M., & Baron, R. J. The holographic hypothesis of memory structure in brain function and perception. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Volume II). San Francisco: W. H. Freeman, 1974.

Pylyshyn, Z. Computation and cognition: Issues in the foundations of cognitive science. *The Behavioural and Brain Sciences,* 1980, *3,* 111-169.

Pysh, J. J. & Weiss, G. M. Exercise during development induces an increase in Purkinje cell dendritic tree size. *Science,* 1979, *206,* 230-232.

Rall, W. & Rinzel, J. Branch input resistance and steady attenuation for input to one branch of a dendritic neuron model. *Biophysics Journal,* 1973, *13,* 648-688.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology,* 1969, *81,* 274-280.

Roberts, L. G. Machine perception of three-dimensional solids. In J. T. Tippett, D. A. Berkowitz, L. C. Clapp, C. J. Koester, & A. Vanderburgh (Eds.), *Optical and electro-optical information processing.* Cambridge: MIT Press, 1965.

Rosenblatt, F. *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms.* Washington, D.C.: Spartan, 1961.

Rosenfeld A., Hummel R. A., & Zucker S. W. Scene labeling by relaxation operations. *I.E.E.E. Transactions on Systems, Man and Cybernetics,* 1976, *SMC-6,* 420-433.

Rumelhart, D. E. & McClelland, J. L. *An interactive activation model of the effect of context in perception: Part 2.* (Tech. Rep. 95). San Diego: University of California, San Diego, Center for Human Information Processing, 1980.

Scheibel, M. E. & Scheibel, A. B. Elementary processes in selected thalmic and cortical subsystems—the structural substrates. In F. O. Schmitt (Ed.), *The neurosciences: second study program.* New York: Rockefeller University Press, 1970.

Selz, O. Die Umgestaltung der Grundanschauungen vom Intellektuellen Geschehen. *Kantstudien,* 1927, *32,* 273–280. Translated as: The revision of the fundamental conceptions of intellectual processes. In J. M. Mandler & G. Mandler (Eds.), *Thinking: From association to Gestalt.* New York: Wiley, 1964.

Shepherd, G. M. *The synaptic organization of the brain* (2nd ed.). New York: Oxford University Press, 1979.

Shirai, Y. A context sensitive line finder for recognition of polyhedra. *Artificial Intelligence,* 1973, *4,* 95–119.

Singer, W. The functional organization of cat striate and parastriate cortex: A correlation between receptive field structure and synaptic connectivity. *Experimental Brain Research, Supp. 1.* O. Creutzfeldt (Ed.) *Afferent and intrinsic organization of laminated structures in the brain.* Berlin: Springer, 1978.

Suga, N. & Jen, P. H. S. Disproportionate tonotopic representation for processing CF-FM sonar signals in the mustache bat auditory complex. *Science,* 1976, *194,* 542–544.

Suga, N. & O'Neill, W. E. Neural axis representing target range in the auditory cortex of the mustache bat. *Science,* 1979, *206,* 351–353.

Sutherland, I. E. & Mead, C. A. Microelectronics and computer science. *Scientific American,* 1977, *237*(3), 210–228.

Szentágothai, J. Specificity versus (quasi-) randomness in cortical connectivity. In M. A. B. Brazier & H. Petsche (Eds.), *Architectonics of the cerebral cortex.* New York: Raven, 1978.

Tusa, R. J., Palmer, L. A. & Rosenquist, A. C. The retinotopic organization of Area 17 (striate cortex) in the cat. *The Journal of Comparative Neurology,* 1978, *177,* 213–235.

Volkmar, F. R. & Greenough, W. T. Rearing complexity affects branching of dendrites in visual cortex of the rat. *Science,* 1972, *176,* 1445–1447.

von Neumann, J. First draft of a report on the EDVAC, June 30, 1945. In B. Randall (Ed.), *The origins of digital computers: Selected papers.* Berlin: Springer, 1975.

Westlake, P. R. The possibilities of neural holographic processes within the Brain. *Kybernetik,* 1970, *7,* 129–153.

Willshaw, D. J. *Models of distributed associative memory.* Unpublished doctoral dissertation, University of Edinburgh, 1971.

Willshaw, D. J., Buneman, O. P. & Longuet-Higgins, H. C. Nonholographic associative memory. *Nature,* 1969, *222,* 960–962.

Winograd, T. *Understanding natural language.* New York: Academic Press, 1972.

Wood, C. C. Variations on a theme by Lashley: Lesion experiments on the neural model of Anderson, Silverstein, Ritz, and Jones. *Psychological Review,* 1978, *85,* 582–591.

Zeki, S. M. Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *Journal of Physiology,* 1978, *277,* 273–290.