BIOLOGICAL ORGANIZATION OF MEMORY*

by

James Anderson

Center for Neural Science and
Psychology Department

and

Leon N Cooper
Center for Neural Science and
Physics Department

Brown University
Providence, Rhode Island 02912

*The work on which this article is based was supported in part by the Ittleson Family Foundation.

Although the properties of individual neurons are relatively well understood, the manner in which large interacting networks of these nerve cells produce mental activity remains almost a complete mystery. This is due in part to the complexity of the central nervous systems of higher animals as well as to the great difficulty of observing these systems without destroying them. But it may also be that processes such as those that result in storage and retrieval of memory are of unusual subtlety, involving small changes in the activities of large numbers of neurons. Finding what such changes occur and where and how memory is stored has proven so difficult that, in a moment of Wagnerian passion, the question has been called 'the Holy Grail of neurobiology.' It is no exaggeration however to conjecture that an understanding of the processes by which an animal stores and retrieves memory might be the key to an understanding of the organization of the central nervous system.

Many ways to store and retrieve information exist: filing cabinets, libraries and computers. But the fact that an animal's memory is held in a living structure and is successfully utilized even though the animal may have no idea of where his memories are stored or how they are ordered, places special requirements on theory. Current computer memories, for example, are made of elements in which yes/no information is recorded and which can be recalled by addressing the location of an element. These computers perform sequences of elementary operations with incredible speed and accuracy, completely beyond the capability of living cells. A basic problem in understanding the organization of memory in a biological system is to understand how a vast quantity of information can be stored and recalled by a system

composed of vulnerable and relatively unreliable elements and with no knowledge of how or where the information has been filed. In what follows we propose a model for the biological organization of memory which can be realized using living cells and which displays properties of association and generalization characteristic of animal mental processes.

Local vs. Distributed Storage

The brain is composed of vast numbers of neurons (10^{10} is the estimate commonly given for humans) held together, fed and cleaned by various supporting structures, blood vessels, and glial (meaning glue) cells. It is thought that the information processing and storage functions of the brain are accomplished by the neurons, the other tissue is occupied primarily with housekeeping. A neuron collects information in the form of electrical potential and currents in its dendrite system from the axon branches of other neurons. There potentials are passively propagated to the cell body, where they are integrated. This integrated potential then determines the firing rate of the cell. The electrochemical spikes that result propagate with minimal degradation over sometimes long distances along the axon trunk to all the axon branches. The information, contained in the frequency of spiking or the number of spikes in a burst, is then communicated chemically across synaptic junctions from the axon terminals to the dendrite branches of other neurons producing potentials in the dendrites; thus the information flow continues.

An immediate question is: How specific are the individual neurons? Do these cells correspond, when they become active to highly specific actions, perceptions, concepts or responses? Or are they simply part of larger structures, so that a single neuron is of importance primarily as a participant in

a complex pattern of nerve cell activity? As is often the case in biological systems, both possibilities seem to be realized depending where or how one looks in the nervous system.

Many invertebrates have nervous systems which are composed of a relatively small number of cells precisely connected together, according to genetic instructions. Certain of these cells seem to control a significant fraction of behavior or respond to a particular, often very specific set of input stimuli. Many of these cells are morphologically identifiable and recur in the same position from animal to animal. In a number of species, many such cells are known where pharmacology, detailed connectivity, and function are understood in great detail. Figure 1 is

Figure 1 about here

a sketch of the abdominal ganglion of the marine gastropod mollusc Aplysia californica which shows some of these identified cells.

Aplysia, in common with many animals, withdraws when touched unexpectedly; the details of this reflex behavior have been worked out by Eric Kandel, Ladislav Tauc and others. Cell L-7, for example, controls a large fragment of the withdrawal of the gill during this reflex.

One of the most interesting results of this work has been the demonstration that the reflex is modifiable in its course and amplitude, and that the reflex will "habituate" if the gill is repeatedly touched. Habituation is more complex behavior than simple fatigue; further, habituation in Aplysia and in higher animals are surprisingly similar.

Important for our consideration of memory, is the clear cut demonstration, in this very simple example of learning, of the localization of the

change involved to the strength of the synaptic junctions coupling the sensory neurons, that becomes active when the animal is touched, and the motor neuron, that, when active, cause the muscles to contract effecting the withdrawal. (Kandel, 1976.)

In invertebrates, we seem to have a system where certain cells correspond to a significant and specific fragment of behavior. Is the same true in higher animals? To a certain extent there is precision of function. Cerebral cortex sometimes displays a surprisingly precise organization. For example, the parts of the cortex concerned with receiving inputs from the visual system map the visual field onto the surface of the cortex in a precise, though spatially distorted, map. The details of connections, for example, the orientation selectivities of the cortical cells, have been shown to be precisel specified, apparently from birth, though they are modifiable to a degree depending on early visual experience. (Lund, 1978.) And many other examples exist. But can such specificity be the rule for all nervous system function?

A leading exponent of the view that a model of the brain can be based on single neuron specificity is Horace Barlow. He has proposed a set of "dogmas" that connect the observed properties of single neurons to psychological function and brain organization, suggesting that "the sensory system is organized so as to achieve as complete a representation of the sensory stimulus as possible with the minimum number of active neurons." "Perception," he states, "corresponds to the activity of a small selection from ... 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." (Barlow, p. 371, 1972.)

Neurons in this type of brain model have been unkindly characterized by critics as "yellow volkswagen detectors" or "grandmother cells," i.e., cells which respond when and only when a yellow volkswagen or the appropriate grandmother appears.

In an opposing point of view, one notes the anatomical homogeneity of large regions of cerebral neocortex. Different areas, though showing important variation from region to region, are basically similar to each other suggesting that one is looking at variations on a subtle, possibly complicated but repeated organizational scheme. Neocortex differs in this regard from those older portions of the brain, for example, those in the brain stem, which seem specially wired for specific purposes.
In addition, the evolution of neocortex has been extremely rapid--an explosive growth of this area of the brain having occured in humans in only a few million years. This suggests that a simple method of organization is repeated over and over by adding more cells and folding this cortical surface to create more area so that the entire structure can be fitted into a skull of reasonable volume. In addition, outside of a few well defined regions, the results of damage to cortex are often diffuse and difficult to describe, and have been observed to depend more on the size of the leison than on its exact location.

Such considerations led Karl Lashley to his proposal that nervous system organization is distributed. He wrote:

> 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 region. (p. 478.)

(Engram is Lashley's term for the physical change corresponding to memory.)

A similar controversy occured in the 19th centry between exponents of cortical localization--the most extreme example is the phrenologist, Gall--and those who favored a more holistic approach to brain organization, such as Goltz, who felt functions were not so rigorously separated.

Actual brain organization no doubt shows both of these aspects. Even in Aplysia, the large identifiable cells may participate in many different kinds of behavior. And in mammalian cerebral cortex, single cells may display astonishing specificity of response, where only very precise combinations of stimuli will induce them to fire.

Binary vs. Analog

Probably the best known and most influential model of the brain is that proposed in 1943 by Warren McCulloch and Walter Pitts. The history of this work demonstrates the practical value of a good model: among other things, the McCulloch-Pitts neuron and the logical notation developed by them influenced John von Neumann when he outlined the architecture of the first modern digital computer. (von Neumann 1945.)

Neurophysiologists of that era were tremendously impressed with the easily visible action potential, the electrochemical cataclysm which transmits information from one end of a neuron to the other. Action potentials are "all or none," that is, they are there or they are not with no intermediate stages.

This lead McCulloch and Pitts to approximate the brain as a set of binary elements which were neurons that were either on or off. They used these binary elements to realize the statements of formal logic. In the abstract

of their 1943 paper they stated: "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.)

We find in the 1943 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 st time quantum reflects the state of the inputs to the elements at the n th time quantum. The main result of their paper was the proof that nets of such neurons were perfectly general in that they could realize any logical expression. A digital computer can be viewed as a machine constructed of McCulloch-Pitts neurons.

Although binary logic was adequate to prove the theorems of the 1943 paper it was not satisfactory as a brain model. One of the most striking aspects of the nervous system is its ability to perform in the presence of noise and with unreliable elements. The nervous system also has the ability to respond to related but differing stimuli and shows a degree of tolerance to perturbations of its inputs, all properties difficult to realize with logical devices, where deviations from perfect accuracy often produce catastrophe. This point was amply clear to McCulloch and Pitts, although not always to those who followed them.

In 1947, Pitts and McCulloch wrote another paper explicitly discussing these problems. With their new more restricted approach, although their assumptions were much more realistic, they could not expect a result of the

generality they produced in 1943. They approximated inputs to the brain as continuous valued distributions--not the discrete 0 and 1 of binary logic--and used transformations and operations on these spatially distributed inputs. They proposed a simple distributed model for the superior colliculus, a midbrain structure which directs the eyes to important points in space. It was known that the projection of the retina to the colliculus forms a precise two-dimensional map of visual space on the surface of the colliculus. (See Figure 2.)

Figure 2 about here

Pitts and McCulloch proposed that the colliculus takes the "weighted center of gravity" of the continuous function describing cell activity on the surfact of the colliculus and directs the gaze to that point.

This model strikingly foreshadows later work on the colliculus with its emphasis on the simultaneous activity of many neurons at once. As we now realize, the colliculus may provide us with one of the best examples we have of a distributed system. Although the retinal efferents that project to the colliculus form a very precise, fine grained map, cells later in the system, physically a few millimeters below the very precise cells, respond to stimuli over a wide area of visual space. Thus we have the apparently paradoxical situation--which seems to be true of other parts of the brain as well--that great precision of response is generated by systems composed of cells which progressively show less and less selectivity as the motor output of the system is approached. (McIlwain, 1976.)

McCulloch was very much aware of the importance of moving away from the binary neuron model. As he stated in one of his last talks, "For our

purpose of proving that a real nervous system could compute any number that a Turing machine could compute with a fixed length of tape, it was possible to treat the neuron as a simple threshold element. Unfortunately, this mislead many into the trap of supposing that threshold logic was all one could obtain in hardware or software. This is false." (McCulloch, p. 393, 1965.)

A Simple Model of a Distributed Memory

Listen once more to Lashley.

> Consideration of the numerical relations of sensory and other cells makes it certain, I believe, that all 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 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 involved, I believe that even the reservation of individual synapses for special associative reactions is impossible. (p. 479-480.)

From this point of view there are no privileged sites in the brain for the storage of memory items in isolation from each other. This seems, at first, very unpromising since individual memories would interfere with each other; but, as we shall see the problem has a solution and memories so constructed can function as well as local memories.

Lashley suggests, and we assume in our distributed models that what is of importance in the operation of the system are activity patterns, simultaneous activities of many different neurons. We define a "trace" as the elementary unit of organization that is processed as a whole and that is large simultaneous spatially distributed pattern of individual neuron activities. It is possible to develop a mathematical structure which lets these

activity patterns take on a life of their own and act very much like primitive entities in their own right. Thus we have moved a step away from individual neuron discharges, which have become small components of the elementary units of nervous activity.

We denote such activity patterns (which will sometimes represent items to be "remembered") by vectors: f^1, f^2, ... f^K [or, more precisely, N-tuples in the N dimensional space composed of N neurons]. In a memory with local storage the individual items would be stored separately. But in a distributed memory they are stored--so to speak--on top of one another. How then can they be distinguished?

To illustrate this basic point we construct a very simple model in which a memory formed by simply taking the vector sum of all these traces, allowing complete interaction of the different inputs

$$s = \sum_{k=1}^{K} f^k$$

Since a component of the vector, s, is the sum of corresponding components of the individual items,

$$s_i = \sum_k f_i^k ,$$

each neuron or synapse (where this memory is presumably held) participates in the storage of all of the individual memory items. By this summation, we have therefore lost information. But much information can be shown to remain.

Suppose an input, f^ℓ, which might have been one of the traces stored in the memory appears. How can it be recognized? We could write the memory, if f^ℓ was contained in it, in the form

$$s = f^{\ell} + (\text{noise})$$

where all the other stored items are called noise. When written this way, we see a problem in signal detection theory: detect the presence or absence of f^{ℓ} in the midst of noise. There are a number of ways to answer this question that have been developed by communications engineers and statisticians.

One simple technique uses the so-called "matched filter" which gives a single number which is large if the input was present in the memory and small or zero if it is not. It is formed as the "dot" or "inner" product of the memory and the input. The inner product of two vectors a and b is defined as

$$a \cdot b = (a, b) = \sum_i a_i b_i$$

Where a_i and b_i are the i th components of a and b respectively.

Filter output with f as input is given by

$$\text{output} = s \cdot f$$

Suppose f^{ℓ} is part of s, i.e., it was stored, then

$$\text{output} = f^{\ell} \cdot f^{\ell} + \sum_{k \neq \ell} f^{k} \cdot f^{\ell}$$

Suppose that the vectors we stored in s are 'distinct' from each other. In particular, let us assume that all the f^{k} stored in s are orthogonal to one another, that is, the inner product $f^{k} \cdot f^{\ell} = 0$ if $k \neq \ell$. (Orthogonal vectors are at right angles to each other in a high dimensional space.) Then we see that all the inner products in the second term above are zero and the output is equal to $f^{\ell} \cdot f^{\ell}$ which is also a positive number. Of course, the

real world could hardly be expected to give rise to such a simple picture. However, if we assume, instead of orthogonality, statistical independence, i.e., the components for the different traces are chosen as different samples from the same probability distribution, the memory will still work, though with a certain amount of noise. This is true because two independent random vectors are orthogonal to each other on the average.

It is possible to calculate a signal to noise ratio for statistically independent traces and show that the signal to noise ratio is (1) directly proportional to the number of elements in the vectors and (2) inversely proportional to the number of stored traces. For memories formed from high dimensionality random vectors lying on the unit hypersphere, computer simulations show that the recognition system rarely makes a mistake when the number of stored traces is less than a few percent of the dimensionality of the vectors. Even when the number of traces approaches half of the dimensionality of the vectors, the system can on the average distinguish stored from non-stored vectors although it makes many errors.

In addition to its ability to distinguish between stored and non-stored vectors such a system is highly parallel in that all interactions between the input and stored elements (the inner product) can be arranged to take place simultaneously, with one overall summation giving the output. It works better, i.e., the signal to noise ratio gets larger, as the dimensionality of the vectors, the number of storage elements, grows larger. Since the brain may contain exceedingly large numbers of storage elements this is a desirable trait. It is also not typical, since most complex information storage and retreival systems--large libraries and modern bureaucracies are good examples--break down if they become too large. Such a

system is relatively damage resistant since loss of a few elements make little difference. In addition, if correlated vectors are stored, the common part of the inputs will tend to be abstracted out and the system will respond to the "average" or "prototype" even though it may never have seen it. This leads to testable predictions. (For a discussion of some psychological aspects of these models see Anderson, 1973, 1977 and Anderson, et al. 1977.)

ASSOCIATIVE MEMORY

Association has been known to be a prominent feature of human memory for over two thousand years. Aristotle observed that "Acts of recollection happen because one change is of a nature to occur after another ... Whenever we recollect, then, we undergo one of the earlier changes until we undergo the one after which the change in question habitually occurs." (de Memoria, 451^b 10-16, translated by Sorabji, 1972.) Association is not a logical process. Associations are formed because of contiguity, happenstance, similarity, or a number of other capricious events external to the observer.

William James presented a modern and mechanistic view of association, hypothesizing that "When two elementary brain processes have been active together or in immediate succession, one of them, in re-occuring, tends to propogate its excitement into the other." (p. 265) and, emphasizing the a-logical nature of association: "It will be observed that the object called up may bear any logical relation whatever to the one which suggested it." (p. 284)

The first modern formal brain models that attempted in an essential way to explain this aspect of memory seem to have been inspired by holograms,

which can be used to associate two images under the appropriate conditions. (See Pribram, Nuwer and Baron, 1974; and Cavanagh, 1972 for the references to this large literature.) Holograms, being distributed, also show the very desirable properties of noise and damage resistance. Although the existence of pure Fourier transform holography in the head seems unlikely, current models that seem promising can be described as generalized holograms, which show the noise and damage resistance of optical holograms. In what follows we present such a model.

Space of Events and Representations

The duration and extent of an "event" should be defined self-consistently by the interaction between the environment and the system itself. We proceed at first, though, as if an event is a well-defined objective happening and envision a space of events E, labeled e^1, e^2, ... e^K. Imagine that these are mapped by the sensory and early processing devices of the system through the mapping, P (processing) into signal distributions in the neuron space, f^1, f^2, ... f^K. The mapping P is denoted by the double arrow in Figure 3. For the moment, we maintain the fiction that this mapping is not modified by experience. What actually seems to be the case is that such early processing systems are at least partially constructed in the youth of the animal and become "hardened" at some relatively early stage in its development.

Although we do not discuss the mapping P in any detail, it can be very complex, and has been optimized for its appropriate functioning in the animal's life in the process of evolution. It must be rich and detailed enough so that enough information is preserved to allow the organism to function. We assume that the mapping P, from E to F, has the fundamental

property of preserving, in a sense not yet completely defined, the closeness or separateness of events.

Figure 3 about here

Two events e^{ν} and e^{μ} map into f^{ν} and f^{μ} whose separation is related to the separation of the origina events. In a vector representation, we imagine that two events as simil ar as a white cat and a grey cat map into vectors which are close to paral el while two events as different as the sound of a bell and the sight of food map into vectors which are close to orthogonal to each other.

Given the signal distribution in F which is the result of an event in E, we imagine that the signal distribution f is mapped onto another set of neurons, or onto the same set, by a mapping, A, denoted by the single arrow in Figure 3. This latter type of mapping is modifiable, and we propose that it is in such mappings that animal memory is contained.

In what follows, we construct an idealized model of a network which incorporates a modifiable mapping and explore some of its properties.

Consider N neurons, 1, 2, ... N each of which has some spontaneous firing rate r_{io}. We can then define an N-tuple, whose components are the difference between the actual firing rate r_j of the j th neuron and the spontaneous firing rate, r_{jo}, that is

$$f_j \equiv r_j - r_{jo}$$

By constructing two such banks of neurons connected to one another, or even by use of a single bank which feeds back on itself, we arrive at a simplified model, as illustrated in Figure 4.

Figure 4 about here

The actual connections between one neuron and the next are complex and may be redundant. We idealize the network by assuming a single ideal synaptic junction which reflects the effect of all the synaptic contacts between neuron j in the F bank and neuron i in the G bank. (See Figure 5.)

Figure 5 about here

Each of the N incoming neurons in F, is connected to each of the N outgoing neurons, in G, by a single ideal junction. We focus our attention on the region above threshold and below saturation and assume that: the firing rate of neuron i in G, g_i, is mapped from the firing rates of all of the neurons, f_j, in F by:

$$g_i = \sum_{j=1} A_{ij} f_j$$

Although most of the results we obtain do not require so strong an assumption, the simplicity of this linear relation makes it useful. In making this approximation for the range above threshold and below saturation we are focussing our attention on firing rates, on time averages of instantaneous signals in a neuron, or possibly a small population of neurons. We are also using the known integrative properties of dendrite branches.

At this point, the basic theoretical entity in neural models has evolved rom the binary, highly specialized McCulloch-Pitts neuron to something very like an analog integrator. Thus the mathematics becomes more like a branch of linear algebra than automata theory.

There is surprisingly strong physiolo ;ical evidence for linearity to be a good approximation within some range of firing rates. Sensory receptors often show impressively linear translations of generator potentials into

firing frequency. (Fuortes, 1971.) Mountcastle has showed linearity of transmission of some aspects of the sensory stimulus from receptors to cortex, once past an initial non-linear transduction. (Mountcastle, 1967.) In the best understood nervous network, the Limulus eye, neurons act as very good linear integrators and this small nervous system can be modeled to high accuracy as a linear system. (See Knight, Toyoda and Dodge, 1970; Ratliff, Knight, Dodge and Hartline 1974.)

The Associative Mapping, Memory and Mental Processes

Animal memory is likely to be distributed and addressed by association. In addition, there need be no clear separation between memory and "logic." We propose that it is in modifiable mappings of the type A that the memory is stored. The mapping A has the properties of a memory that is non-local, content addressable, and in which "logic" is a result of association and an outcome of the nature of memory itself.

In agreement with such experimental evidence as Kandel's findings for Aplysia and in agreement with the opinion of most neuroscientists, we assume that the physical locus of memory is at the synapses coupling cells and that precisely specified changes in these coupling elements store permanent memory. [They may also account for less permanent memory.]

How might a mapping be put into the network? A synaptic modification scheme apparently first suggested by D. O. Hebb seems the most promising. Hebb's suggestion was stated in a well known passage from the book "Organization of Behavior" as:

> 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. (Hebb, p. 62, 1949.)

A particularly attractive candidate for learning in cerebral cortex are changes in the specialized processes--dendritic spines--which emerge in great numbers from the dendrites of pyramidal cells. At the end of these fine processes, are located synapses. Virtually all synapses on dendrites in pyramids are on spines. Since the dendrite may be quite thick up to the point where the thin spine process emerges, there is a low resistance pathway from the spike initiating region of the cell to the spine. Only microns away, on the other end of the spine is the pre-synaptic contact, thus pre- and post-synaptic cell are in close apposition and small physical changes in spine geometry could produce large changes in synaptic efficacy. Since changes in spine morphology have been demonstrated in response to environmental modification in several contexts, this is a speculative but interesting candidate for the site of at least some kinds of learning. (See discussion and references in pgs. 80-86 of Peters, Palay and Webster, 1976, where Figure 6 was taken.

Figure 6 about here

The synaptic modification assumption above gives a mapping A after the system has had various sets of activity patterns in the F bank, f^{ν} and in the G bank, g^{μ}. We can write

$$A = \sum_{\mu\nu} c_{\mu\nu} g^{\mu} \times f^{\nu}$$

Here x denotes the "outer" product and yields a matrix. Although this is a transparent mathematical form, its meaning as a mapping among neurons deserves some discussion. The ij th element of A gives the strength of the ideal junction between the incoming neuron j in the F bank and the outgoing neuron i in the G bank.

According to this, cells will tend to become correlated in their discharges; a synapse acting in this way is sometimes called a correlational synapse. As formulated by Hebb, the model was not immediately suitable for much mathematical development. However, in the past few years, a variety of brain models have been put forth more or less independently by a number of investigators, which incorporate centrally a learning postulate somewhat like Hebb's. Examples are Willshaw, Buneman and Longuet-Higgins, 1969; Nass and Cooper 1975; Anderson, 1970, 1972; Amari, 1972; Grossberg, 1971; Kohonen, 1972; Cooper, 1974; Little and Shaw, 1975; Wilson, 1975; among others. Kohonen (1977) has reviewed much of this work in his recent book "Associative Memory: A System Theoretic Approach."

Suppose activity pattern f^{ν} is the F bank and activity pattern g^{ν} is in the G bank. We add to the elements A_{ij} increments of the following type:

$$\delta A_{ij} \sim g_i f_i$$

This is proportional to the product of the differences between the actual and the spontaneous firing rates in the pre- and post-synaptic neurons j and i.

For such modifications to occur, there must be a means of communication between the region where the action potentials are initiated and the synapse. Possibilities for this are many: One might be electrotonic conduction from the spike initiating region. There are also a multitude of substrates for the change: changes in membrane resistivity, change in size of the synapse, changes in amount of transmitter, growth of new dendrites or synapses, increase in sensitivity of the junction, and many others. The problem is not too few candidates for modifiability, but too many.

Thus, if only the j^{th} component of the incoming activity pattern, f_j, is non-zero

$$g_i = A_{ij} f_j \ .$$

Since

$$A_{ij} = \sum_{\mu\nu} c_{\mu\nu} g_i^{\mu} f_j^{\nu}$$

the ij th junction strength is composed of a sum of the entire experience of the system as reflected in firing rates of the neurons connected to his junction. Each experience or association ($\mu\nu$), however, is stored over the entire array of N x N junctions. This is the essential meaning of a distributed memory: Each event is stored over a large portion of the system, while at any local point, many events are superimposed.

Recognition and Recollection

The fundamental problem posed by a distributed memory is the address and accuracy of recall of the stored events. Consider first the "diagonal" portion of A,

$$(A)_{\text{diagonal}} \equiv \mathcal{R} \equiv \sum_{\nu} c_{\nu\nu} g^{\nu} \times f^{\nu}$$

An arbitrary event, e, mapped into the signal, f, will generate the response in G

$$g = Af$$

If we equate recognition with the strength of this response, say the inner product

$$(g, g) \ ,$$

then the mapping A will distinguish between those events it contains, the f^{ν}, $\nu = 1, 2, \ldots K$ and other events separated from these. (This is a

slightly different, but related, definition of recognition from that proposed for the summed vector model of a distributed system discussed previously.)

The word "separated" in the above context requires definition. Suppose the vector f^ν are thought to be independent of each other, and to satisfy the requirements that, on the average

$$\sum_{i=1}^{N} f_i^\nu = 0$$

$$\sum_{i=1}^{N} (f_i^\nu)^2 = 1 \ .$$

Any two such vectors have components which are random with respect to one another so that a new vector, f, presented to the F bank above gives a noise like response in the G bank since on the average (f^ν, f) is small. The presentation of a vector seem previously, f^λ, however, gives the response in the G bank

$$f^\lambda = c_{\lambda\lambda} g^\lambda + \text{noise}$$

It can then be shown that if the number of imprinted events, K, is small compared to the dimensionality, N, the signal to noise ratios are reasonable, as discussed earlier.

If we define separated events as those which map into orthogonal vectors, then clearly a recognition matrix composed of K orthogonal vectors $f^1, f^2, \ldots f^K$

$$\mathcal{R} = \sum_{\nu=1}^{k} c_{\nu\nu} g^\nu \times f^\nu$$

will distinguish between those vectors contained and all vectors separated from (perpendicular to) these. Further, the response of the system to a vector

previously recorded is unique and completely accurate

$$f^{\lambda} = c_{\lambda\lambda} g^{\lambda}$$

In this special situation, the distributed memory is as precise as a localized memory.

In addition, this type of memory has the interesting property of recalling an entire associated vector g^{λ} even if only part of f^{λ} is presented. Let

$$f^{\lambda} = f_1^{\lambda} + f_2^{\lambda}$$

If only part of f^{λ}, say f_1^{λ} is presented, we obtain

$$f_1^{\lambda} = c_{\lambda\lambda}(f_1^{\lambda}, f_1^{\lambda})\ g^{\lambda} + \text{noise}$$

The result is the entire response to the full f^{λ} with a reduced coefficient plus noise.

Association

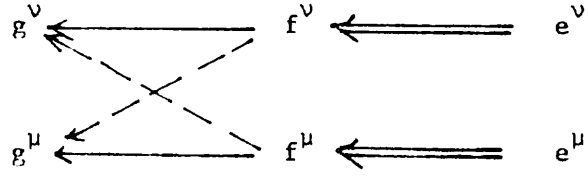
If we now take the point of view that presentation of the events e^{ν} which generates the vector f^{ν} is recollected if

$$f^{\nu} = c_{\nu\nu} + \text{noise}$$

Then the off-diagonal terms

$$\mathcal{A} \equiv \sum_{\mu \neq \nu} c_{\mu\nu} g^{\mu} \times f^{\nu}$$

may be interpreted as containing associations between events initially separated from one another.



where $(f^{\nu}, f^{\mu}) = 0$.

For such terms the presentation of event e^{ν} will generate not only g^{ν} (which is equivalent to the recollection of e^{ν}) but also, and perhaps more weakly, g^{μ} which should result with the presentation of e^{μ}. Thus, for example if g^{μ} will initiate some response, originally a response to e^{μ}, the presentation of e^{ν} when $c_{\mu\nu} \neq 0$ will also initiate this response.

We can thus divide the association matrix A into two parts:

$$A = \sum_{\mu\nu} c_{\mu\nu} g^{\mu} \times f^{\nu} = \mathcal{R} + \mathcal{A}$$

where

$$\mathcal{R} = (A)_{\text{diagonal}} \equiv \sum_{\nu} c_{\nu\nu} g^{\nu} \times f^{\nu} \qquad \text{[recognition and recollection]}$$

and

$$\mathcal{A} = (A)_{\text{off-diagonal}} \equiv \sum_{\mu \neq \nu} c_{\mu\nu} g^{\mu} \times f^{\nu} \qquad \text{[association]}$$

The $c_{\mu\nu}$ are then the direct recollection and association coefficients.

Biological "Logic"

In actual experience, the events to which the system is exposed are not in general highly separated nor are they independent in a statistical sense. There is no reason, therefore, to except that all vectors, f^{ν}, printed into A would be orthogonal or even very far from one another. Rather it seems likely that often large numbers of these vectors would lie close to one another. Under these circumstances, a distributed memory might be "confused" in the sense that it will respond to new events as if they were old, if the new event is close to an old one. It will "recogn ze" and "associate" events never, in fact, seen or associated before.

The memory will tend to categorize stimuli on the basis of the past history of the system. For example, suppose a number of vectors in the memory are of the form

$$f^{\nu} = f^{0} + n^{\nu}$$

where n^{ν} varies randomly, f^{0} will eventually be recognized more strongly than any particular f^{ν} actually presented.

This is a testable prediction; something very much like this seems to occur in a few contexts where it can be checked. (See Anderson, 1977, Section IV for a discussion of the psychological experiments bearing on this point.)

We have here an explicit realization of what might loosely be called biological "logic" which, of course, is not logic at all. Rather what occurs might be described as the result of a built in tendency to "leap to conclusions."

This property has certain similarities to the psychological properties called "generalization" or "abstraction." In these models, generalization grows from the loss of detail of individual instances. Thus generality is gained at the price of precision, a kind of trade-off that seems characteristic of distributed systems.

The system takes the step, to give one example, from cat^{1}, cat^{2}, cat^{3} ... to the general: cat. How fast this step is taken depends on the system's parameters. By altering these parameters it is possible to construct mappings which vary from those which retain the particulars to which they are exposed, to those which lose the particulars and retain only common elements--the central vector of a class.

In addition to errors of recognition, the associative memory also makes errors of association. If, for example, all (or many) of the vectors of a class $\{f^\alpha\}$ associates some particular g^β so the mappings A contains terms of the form

$$\sum_{\alpha=1}^{K} c_{\beta\alpha}\, g^\beta \times f^\alpha$$

with $c_\beta^\alpha \neq 0$ over much of $\alpha = 1, 2, \ldots K$, then the new event e^{K+1} which maps into f^{K+1} as the previous example will not only be recognized (that is the inner product (Af^{K+1}, Af^{K+1}) will be large), but will also associate

$$Af^{K+1} = cg^\beta + \ldots$$

as strongly as any of the vectors in $\{f^\alpha\}$.

If errors of recognition lead to the process described in language as going from particulars to the general, errors of association might be described as going from particulars to a universal: cat^1 meows, cat^2 meows, ... --all cats meow.

There is, of course, no "justification" for this process. Whatever efficacy it has will depend on the structure of the world in which the animal system finds itself. If the world is properly ordered, an animal which "jumps to conclusions" may be better able to react and to adapt to the hazards of its environment. The animal philosopher sophisticated enough to argue "the tiger ate my friend but that does not allow me to conclude that he might want to eat me" could be a recent development whose survival depends on other less sophisticated animals who jump to conclusions.

By a sequence of mappings of the form above, or by feeding the output of A back onto itself, one obtains a fabric of events and connections which is rich as well as suggestive. One easily sees the possibility of a flow of electrical activity influenced both by internal mappings of the form A and the external input. This flow is governed not only by the direct associations $c_{\mu\nu}$ but also by indirect associations due to the overlapping of mapped events. One can easily imagine situations arising in which direct access to an event, or class of events, has been lost while the existence of this event or class of events in A influences the flow of electrical activity.

Self-Organization of the Associative Memory

To make the modification that we have assumed,

$$\delta A_{ij} \sim g_i f_j$$

by any of the mechanisms that might exist, the system must have the activity pattern f^ν in its F bank and g^μ in its G bank. It is easy to obtain f^ν since this is mapped in from the event e^ν by P. But to get g^μ in the G bank may be more difficult in some circumstances since this is what the system is trying to learn.

In what we denote as "active learning," which has been explored extensively in the past, the system is presented with some f^λ, searches for a response and is given some indication of when it is coming closer. When by some procedure or other it finds the right response, say g^ω, it "prints" into A the information,

$$\delta A_{ij} = \eta g_i^\omega f_j^\lambda$$

This information is available at the synaptic junctions at the time of the "print" order since at that time the system is mapping f^{λ}, responding g^{ω}, and thus had just the appropriate activity in the F and G banks. Active learning seems most suitable for a case where a system response to an input is matched against an expected or desired response and judged correct or incorrect. Clearly, much important learning is of this type.

However, there is a type of learning which may not require a search procedure of this kind. It is a type of learning in which an animal is placed in an environment and seems to learn to recognize and to recollect in a passive manner.

One form of a passive learning algorithm (Cooper, 1974) utilizes a distinction between forming an internal representation of events in the external world as opposed to producing a response to these events which is matched against what is expected or desired in the external world.

The simple but important idea is that the internal electrical activity which in one mind signals the presence of an external event is not necessarily, or even likely to be, the same electrical activity which signals the presence of that same event for another mind. There is nothing that requires that the same external event be mapped into the same neural patterns by different animals. What is required for eventual agreement between minds in their description of the external world is not that the electrical signals mapped be identical but rather than the relation of the signals to each other and to events in the external world be the same.

Passive Learning

Call A(t) be the A matrix after the presentation of t events. We write

$$A(t) = \gamma A(t - 1) + \eta g(t) \times f(t)$$

In the equation γ is dimensionless and is a measure of the uniform decay of information at every site, a type of forgetting. One would expect that γ would take values between zero and one. It seems from simulations and analysis that values of γ close to one are of most interest (i.e., forgetting is small). For convenience we normalize all input vectors so that $(f, f) = 1$.

If we now say that $g(t)$ is

$$g(t) = \gamma A(t - 1)\ f(t) + g_R(t) + g_A(t) \ .$$

We see that the total post-synaptic potential events are composed of three terms: a passive response, $\gamma A(t - 1)\ f(t)$, an active but random term, $g_R(t)$, and an active response, $g_A(t)$. For purely passive learning, we consider only the first term so that

$$\delta A = \eta g(t) \times f(t) = \gamma \eta A(t - 1)\ f(t) \times f(t) \ .$$

Here the post-synaptic potentials are just those produced by the decayed existing mapping, $\gamma A(t - 1)$ when the vector $f(t)$ in F is mapped into G

$$g(t) = \gamma A(t - 1)\ f(t)$$

The passive learning algorithm is then

$$A(t) = \gamma A(t - 1)\ [1 + \eta f(t) \times f(t)]$$

where η is presumably much smaller than one. Before any external events have been presented, A has the form $A(0)$ which could be random or could contain information which has been programmed genetically. It also will contain the connectivity of the network.

With this algorithm, after K events, e^1, e^2, ... e^K which map into f^1, f^2, ... f^K, A has the form

$$A(K) = \gamma^K A(0) \prod_{\nu=1}^{K}{}_{o} (1 + \eta f^\nu \times f^\nu)$$

where Π_o is an ordered product in which the factors with lower indices stand to the left.

It is striking that the passive learning algorithm generates its own response $A(0)f^\nu$ to the incoming vector f^ν, a response that depends on the original configuration of the network through $A(0)$ and on the vector f^ν mapped from the event e^ν. For example, if f^ν is the only vector presented, A eventually takes the form

$$A \sim g^\nu \times f^\nu$$

where

$$g^\nu \equiv A(0)f^\nu .$$

Special Cases of A

To illustrate some of the properties of passive learning, consider four special cases of interest; in all of these, η is assumed to be constant and small.

(1) If the K vectors are orthogonal, A becomes

$$A(K) = \gamma^K A(0) \left(1 + \eta \sum_{\nu=1}^{K} f^\nu \times f^\nu\right)$$

Letting $A(0)\, f^\nu \equiv g^\nu$, the second term takes the form of the diagonal part of A,

$$(A)_{\text{diagonal}} \equiv \mathcal{R} = \eta \sum_{\nu=1}^{K} g^\nu \times f^\nu,$$

and will serve for recognition of the vectors f^1 ... f^K. It should be observed that the associated vector g^ν are not given in advance, they are

generated by the network. If η is small, however, this seems inadequate for recognition, since the recognition term will be weak. One would expect recognition to build up only after repeated exposure to the same event.

(2) The passive learning algorithm does build up recognition coefficients for repeated inputs of the same events. If f^0 is presented ℓ times, A becomes

$$A(\ell) \simeq \gamma^{\ell} A(0) \ (1 + e^{\ell\eta} f^0 \times f^0) \ .$$

If ℓ is large enough so that $e^{\ell\eta} > 1$, the recognition term will eventually dominate.

(3) The presentation of orthogonal vectors $\ell_1, \ell_2, \ldots \ell_m$ times results in a simple generalization of the second result. When $\gamma = 1$, for simplicity,

$$A(\ell_1 + \ell_2 + \ldots + \ell_m) = A(0) \ (1 + \sum_{\nu=1}^{m} e^{\ell_\nu \eta} \ f^\nu \times f^\nu)$$

which is just a separated recognition and recall matrix

$$A \simeq \sum_{\nu=1}^{m} c_{\nu\nu} g^\nu \times f^\nu$$

if

$$e^{\ell_\nu \eta} \equiv c_{\nu\nu} >> 1 \ .$$

(4) Some of the effects of non-orthogonality can be seen by calculating the result of an input consisting of ℓ noisy vectors distributed around a central f^0

$$f^\nu = f^0 + n^\nu$$

Here n^ν is a stochastic vector whose magnitude is small compared to that of f^0. We obtain

$$A(\ell) \simeq \gamma^\ell A(0) \exp(\ell\eta \frac{n^2}{N}) (1 + e^{\ell\eta} f^0 \times f^0)$$

where n is the average magnitude of n^ν. We see that the generated $A(\ell)$ with the additional factor due to the noise is of the form for recognition of f^0. Thus the repeated application of a noisy vector of the form above results in an A which recognizes the central vector f^0.

Association Terms

Off-diagonal or associative terms can be generated as follows. Assume that A has attained the form

$$A = \sum_{\nu=1}^{K} A(0) f^\nu \times f^\nu = \sum_{\nu=1}^{K} g^\nu \times f^\nu .$$

Now present the events e^α and e^β so they are associated and the vectors f^α and f^β map together. The precise conditions which result in such a simultaneous mapping of f^α and f^β will depend on the construction of the system. The simplest situation to imagine is that in which $(f^\alpha + f^\beta)/\sqrt{2}$ is mapped if e^α and e^β are presented to the system close enough to each other in time. We may assume that e^α and e^β are separated so that $(f^\alpha, f^\beta) = 0$. In the F bank of neurons we then have $(f^\alpha + f^\beta)/\sqrt{2}$ where the vector is normalized for convenience.

After one such presentation of e^α and e^β, A becomes, with $\gamma = 1$,

$$A(1) = \sum_{\nu=1}^{K} g^\nu \times f^\nu + \frac{\eta}{2} (g^\beta \times f^\alpha + g^\alpha \times f^\beta) .$$

The second term gives the association between α and β with the coefficient

$$c_{\alpha\beta} = c_{\beta\alpha} = \eta/2$$

which presumably, except in special circumstances, would be small. If f^{α} and f^{β} do not occur again in association the coefficients $c_{\alpha\beta}$ and $c_{\beta\alpha}$ remain small compared to $c_{\alpha\alpha}$ and $c_{\beta\beta}$. However if $(f^{\alpha} + f^{\beta})/2$ is a frequent occurance, appearing, for example, ℓ times, the coefficient of the cross term becomes

$$c_{\alpha\beta} \approx \frac{\ell\eta}{2}$$

as large as the recognition coefficient.

Structure of the Mapped Space

In order that the mapped spaces be useful to the animal in forming an internal representation of the external world, as well as for eventual concordance between animals in their description of the external world, the relations in the mapped spaces must in some sense be in correspondence to those in the external world. We show here how this comes about in one simple case. Assume that in the external world there are K separated events e^1, e^2, ... e^k which map into f^1, f^2, ... f^k which are orthogonal to one another. Under these circumstances the eventual form of A upon repeated presentation of the e's will be

$$A = \sum_{\nu=1}^{K} g^{\nu} \times f^{\nu} .$$

There is no need that the g^{ν} and f^{ν} for one animal which are mapped from e^{ν} be the same as the g'^{ν} and f'^{ν} mapped from e^{ν} by another animal. What is required is that the "structure" of the two mappings be "similar." To make this last sentence precise in a manner stronger than is actually re-

quired, we ask that for every e^{α} there exists an f^{α} and g^{ι} such that the inner products

$$(g^{\mu}, g^{\nu}) = (f^{\mu}, f^{\nu})$$

This requirement will be met if A(0 is unitary, that is

$$\sum_{i=1}^{N} A_{ij}(0)\, A_{ik}(0) \simeq \delta_{jk}$$

which is a requirement on the original connectivity of the network.

This can easily be arranged. For example if we choose

$$A(0) = I$$

the above requirement will be satisfied. Even a random A(0) with evenly distributed positive and negative entries will, on the average satisfy this requirement and therefore result in a mapped space with the same communities and classes as the original event space.

If we combine this result with the prior results on the build-up of association coefficients, we see that (at least for separated inputs)

(1) The classes or communities of the mapped space, G, are the same as those of the external or input spaces, E and F.

(2) Classes or events which are associated in the external space, those which occur in association during a learning period, become associated in the mapped spaces so that, after the learning period the occurrence of one member of the associated classes or events in the external space, E, and therefore in the input space, F, will map both members of the associated classes or events in G even though they are very different types of events.

CONCLUSION

Although we have emphasized how associative memories can organize themselves by interaction with the environment, nothing in this approach limits the possibility of genetically determined connectivity, pre-wired feature detectors or other innate organization. Obviously not all synaptic junctions need be or are likely to be modifiable. In any actual system pre-wired and non-modifiable junctions will function together with those that are modifiable. [One way pre-wiring can be put into the associative matrix, A, is as its initial value, A(0). Non-modifiable synapses can be treated along with those that are modifiable, for example by dividing A into two parts one modifiable, the other not.]

A most exciting result is the demonstration that a system can passively construct a memory by interaction with its environment in such a way as to learn the associations present in its environment. This is accomplished without explicit prior instruction about the environment the system will encounter. This result is especially important since it seems clear that any theory of the central nervous system must account for that system's capacity to function in widely different situations that are not likely to have been pre-programmed in any but the most general fashion.

As a next step, it is important to confront the various assumptions and theoretical constructions with experiment. This is difficult since we are looking for subtle changes in the behavior of living cells which are not easy to observe. Yet some progress has been made. In Aplysia, for example, synaptic modification has been directly observed. And it is possible that visual cortex, where much experimental and theoretical work has been done, might prove to be a region in which theoretical ideas can be confronted

by experiment.

Synaptic modification dependent on inputs alone, of the type directly observed in Aplysia, is already sufficient to construct the simplest memory given as our first example

$$s = \sum_k f_i^k$$

a memory that distinguishes what has been seen from what has not, but does not easily separate one input from another. To distinguish between inputs as well requires synaptic modification dependent on input and output (or more generally, dependent on information that exists at different places on the neuron membrane). In order that such modification take place, the information must be communicated from, for example, the axon hillock to the synaptic junction to be modified. This implies the possibility of internal communication of information within the neuron. If a mechanism for such communication exists one might guess that specific forms evolved in various ways and that various types of two (or higher) point modification exist.

It is tempting to conjecture that a liberating evolutionary step was just the development of this means of internal communication which, coupled with the ability of synapses to modify, created the possibility for a new organizational principle.

One must also show that it is possible, at least in principle, to construct systems, using networks such as those described above, that can accomplish at least some of the tasks done so easily by the educated animal. Here psychological models may be of great value. However, much remains to be done. Even such a simple seeming question as how a common pattern is recognized remains unanswered.

It seems likely, therefore, that some fundamental new ideas remain to be introduced before we can say that we have arrived at an understanding of the relation between brain function and brain as a biological system--in the words of William James "...the scientific achievement before which all past achievements would pale."

REFERENCES

Amari, S., "Learning Patterns and Pattern Sequences by Self Organizing Nets of Threshold Elements," IEEE Transactions on Computers, C-21, 1197-1206 (1972).

Anderson, J. A., "Two Models for Memory Organization Using Interacting Traces," Mathematical Biosciences, 8, 137-160 (1970).

Anderson, J. A., "A Simple Neural Network Generating an Interactive Memory," Mathematical Biosciences, 14, 197-220 (1972).

Anderson, J. A., "A Theory for the Recognition of Items from Short Memorized Lists," Psychological Review, 80, 417-438 (1973).

Anderson, J. A., "Neural Models with Cognitive Implications," in D. LaBerge and S. J. Samuels (Eds.), Basic Processes in Reading: Perception and Comprehension, Hillsdale, New Jersey: Erlbaum Associates (1977).

Anderson, J. A., Silverstein, J. W., Ritz, S. A. and Jones, R. S., "Distinctive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model," Psychological Review, 84, 413-451 (1977).

Barlow, H. B., "Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology," Perception, 1, 371-394 (1972).

Cavanagh, P., "Holographic Processes Realizable in the Neural Realm: Prediction of Short Term Memory Performance," Doctoral Disseration: Carnegie-Mellon University (1972), Dissertation Abstracts International, 33, 3280B (1973).

Cooper, L. N, "A Possible Organization of Animal Memory and Learning," in B. Lundquist and S. Lundquist (Eds.), Proceedings of the Nobel Symposium on Collective Properties of Physical Systems, New York: Academic Press (1974).

Fuortes, M. G. F., "Generation of Responses in Receptors," in W. R. Lowenstein (Ed.), Handbook of Sensory Physiology, I, Principles of Receptor Physiology, Berlin: Springer (1971).

Grossberg, S., "Pavlovian Pattern Learning by Nonlinear Neural Networks," Proceedings of the National Academy of Sciences, 68, 828-831 (1971).

Hebb, D. O., "The Organization of Behavior," New York: Wiley (1949).

James, W., "Psychology: Briefer Course," New York: Collier (1962). (Original publication: 1890).

Kandel, E. R., "Cellular Basis of Behavior: An Introduction to Behavioral Neurobiology," San Francisco: W. H. Freeman (1976).

Knight, B. W., Toyoda, J. I., and Dodge, F. A., Jr., "A Quantitative Description of the Dynamics of Excitation and Inhibition in the Eye of Limulus," The

Journal of General Physiology, 56, 421-437 (1970).

Kohonen, T., "Correlation Matrix Memories," IEEE Transactions on Computers, C-21, 353-359 (1972).

Kohonen, T., "Associative Memory: A System Theoretic Approach," Berlin: Springer-Verlag (1977).

Lashley, K. S., "In Search of the Engram," Symposia of the Society of Experimental Biologists, No. 4, Physiological Mechanisms in Animal Behavior, New York: Academic Press (1950).

Little, W. A. and Shaw, G. L., "A Statistical Theory of Short and Long Term Memory," Behavioral Biology, 14, 115-133 (1975).

Lund, R. D., "Development and Plasticity of the Brain," New York: Oxford (1978).

McCulloch, W. S., "Embodiments of Mind," Cambridge, Massachusetts: MIT Press (1965).

McCulloch, W. S. and Pitts, W. H., "A Logical Calculus of Ideas Immanent in Nervous Activity," Bulletin of Mathematical Biophysics, 5, 115-133 (1943).

McIlwain, J. T., "Large Receptive Fields and Spatial Transformations in the Visual System," In R. Porter (Ed.), International Review of Physiology: Neurophysiology II, Volume 10, Baltimore: University Park Press (1976).

Mountcastle, V. B., "The Problem of Sensing and the Neural Coding of Sensory Events," In G. C. Quarton, T. Melnechuk and F. O. Schmitt (Eds.), The Neurosciences, New York: Rockefeller University Press (1967).

Nass, M. M. and Cooper, L. N, "A Theory for the Development of Feature Detecting Cells in Visual Cortex," Biological Cybernetics, 19, 1-18 (1975).

Peters, A., Palay, S. L., and Webster, H. deF, "The Fine Structure of the Nervous System: The Neurons and Supporting Cells," Philadelphia: W. B. Saunders (1976).

Pitts, W. H. and McCulloch, W. S., "How We Know Universals: The Perception of Auditory and Visual Forms," Bulletin of Mathematical Biophysics, 9, 127-147 (1947).

Pribram, K., Nuwer, M., and Baron, R., "The Holographic Hypothesis of Memory Structure in Brain Function and Perception," in D. H. Krantz, R. C. Atkinson, R. D. Luce and P. Suppes (Eds.), Contemporary Developments in Mathematical Psychology, Volume II, San Francisco: W. H. Freeman (1974).

Ratliff, F., Knight, B. W., Dodge, F. A., Jr., and Hartline, H. K., "Fourier Analysis of Dynamics of Excitation and Inhibition in the Eye of Limulus Amplitude, Phase, and Distance," Vision Research, 14, 1155-1168 (1974).

Sorabji, R., "Aristotle on Memory," Providence, Rhode Island: Brown University Press (1972).

von Neumann, J., "First Draft of a Report on the EDVAC," (dated June 30, 1945), Moore School of Electrical Engineering Technical Report, University of Pennsylvania, reprinted in: The Origins of Digital Computers, 2nd ed., edited by B. Randall, Berlin: Springer (1975).

Willshaw, D. J., Buneman, O. P., and Longuet-Higgins, H. C., "Non-Holographic Associative Memory," Nature, 222, 960-982 (1969).

Wilson, H. R., "A Synaptic Model for Spatial Frequency Adaptation," Journal of Theotical Biology, 50, 327-352 (1975).

FIGURE CAPTIONS

Figure 1 -- "Map of identified cells in the abdominal ganglion of Aplysia californica indicating the most common positions of the identified cells. The identified cells are labeled L or R (left or right hemiganglion) and assigned a number. The hemiganglia are arbitrarily subdivided into quarter-ganglia. Cells that share similar properties generally have similar labels, e.g., $L9_{G1}$ and $L9_{G2}$, and L14A, L14B, and L14C. Cells that are members of clusters are identified by the cluster name and a subscript identifying the behavioral function of the cell, e.g., LD_{H11} and LD_{H12}, two heart inhibitors belonging to the LD cluster." [From E. Kandel, pg. 226 (1976).]

Figure 2 -- Simple distributed computational model of superior colliculus. A calculation of the "weighted center of gravity" of the distribution of afferent impulses on the surface of the colliculus (a distorted two-dimensional map of visual space) directs eyes to that point. [From Pitts and McCulloch, pg. 127 (1947) Figure 6.]

Figure 3 -- The N neurons in the F bank are connected via synaptic junctions to the N neurons of the G bank. [From Cooper, pg. 254 (1974).]

Figure 4 -- The ideal associator unit. Each of the N incoming neurons in F is connected to each of the N outgoing neurons in G by a single ideal junction. (Only the connections to i are drawn.) We assume that the firing rate of neuron i in G, g_i, is mapped from the firing rates of all of the neurons in F by: $g_i = \sum_j A_{ij} f_j$. [From Cooper, pg. 254 (1974).]

Figure 5 -- The ideal junction. [From Cooper, pg. 255 (1974).]

Figure 6 -- An electron micrograph of a dendritic spine (magnification about 40,000) in a freeze-fracture preparation of the cerebellum of an adult rat. Although this is not a cell from the cerebral cortex, spine anatomy is similar to cortical spines in most respects. [From Peters, Palay and Webster, pg. 84, (1976).]

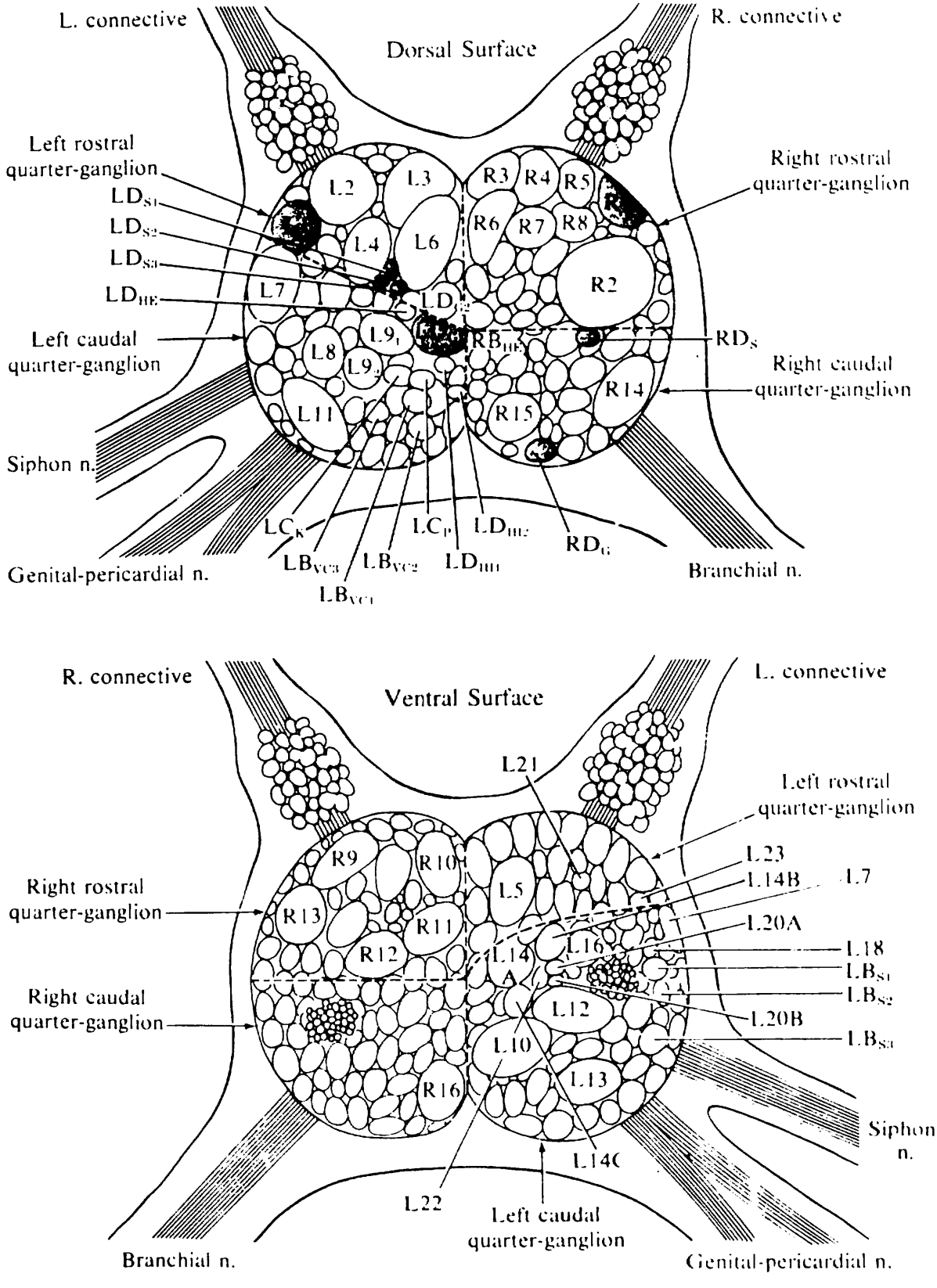

L. connective
Dorsal Surface
R. connective
Left rostral quarter-ganglion
Right rostral quarter-ganglion
Left caudal quarter-ganglion
Right caudal quarter-ganglion
Siphon n.
Genital-pericardial n.
Branchial n.
L2
L3
L4
L6
L7
L8
L11
R2
R3
R4
R5
R6
R7
R8
R14
R15
Ventral Surface
L21
L23
L14B
L7
L20A
L18
L12
L20B
L10
L13
L14C
L22
R9
R10
L5
R13
R11
R12
L14A
L16
R16

FIGURE 1

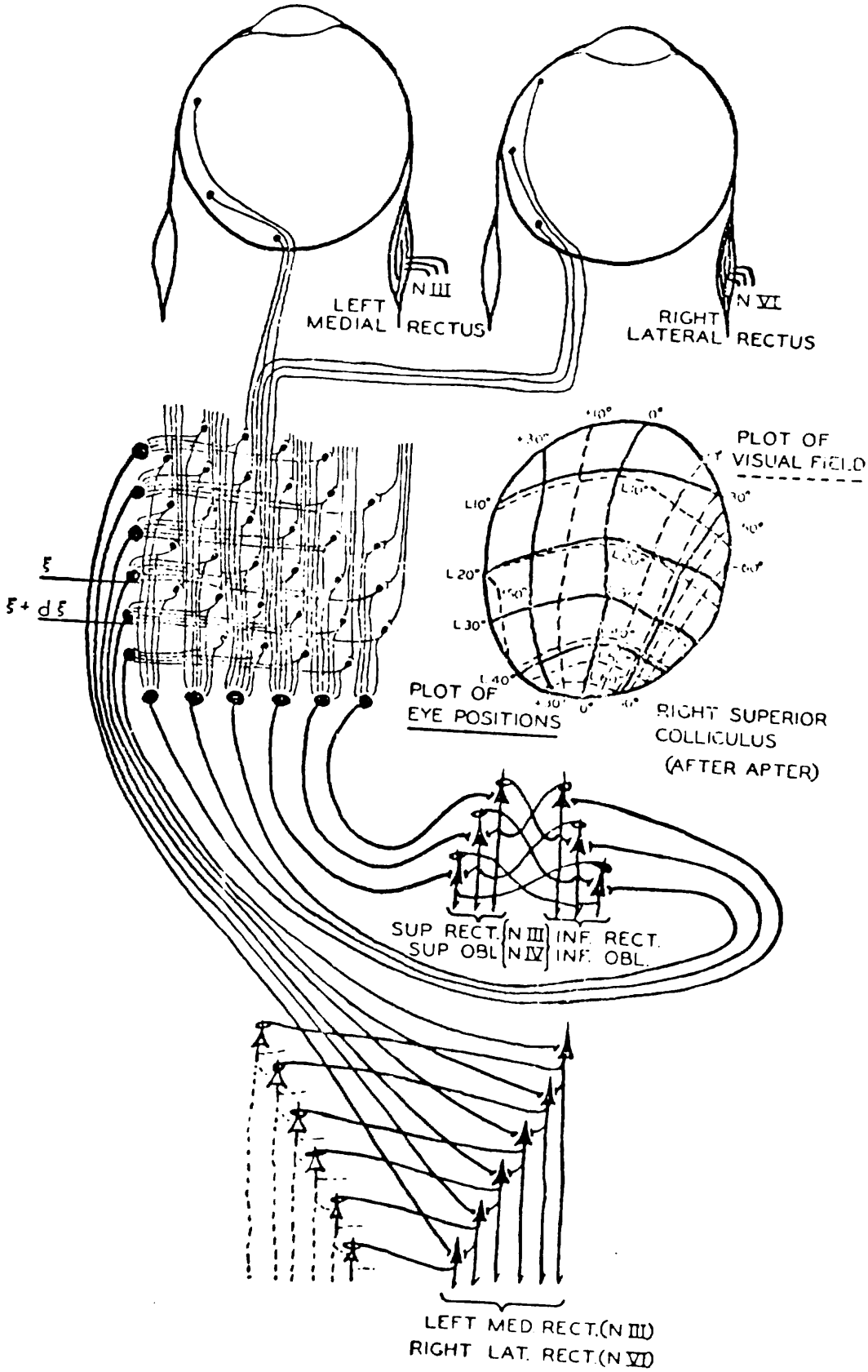

LEFT MEDIAL RECTUS
N III
RIGHT LATERAL RECTUS
N VI
PLOT OF VISUAL FIELD
PLOT OF EYE POSITIONS
RIGHT SUPERIOR COLLICULUS (AFTER APTER)
ξ
ξ + dξ
SUP RECT. N III INF. RECT.
SUP OBL N IV INF. OBL.
LEFT MED RECT.(N III)
RIGHT LAT. RECT.(N VI)

FIGURE 2

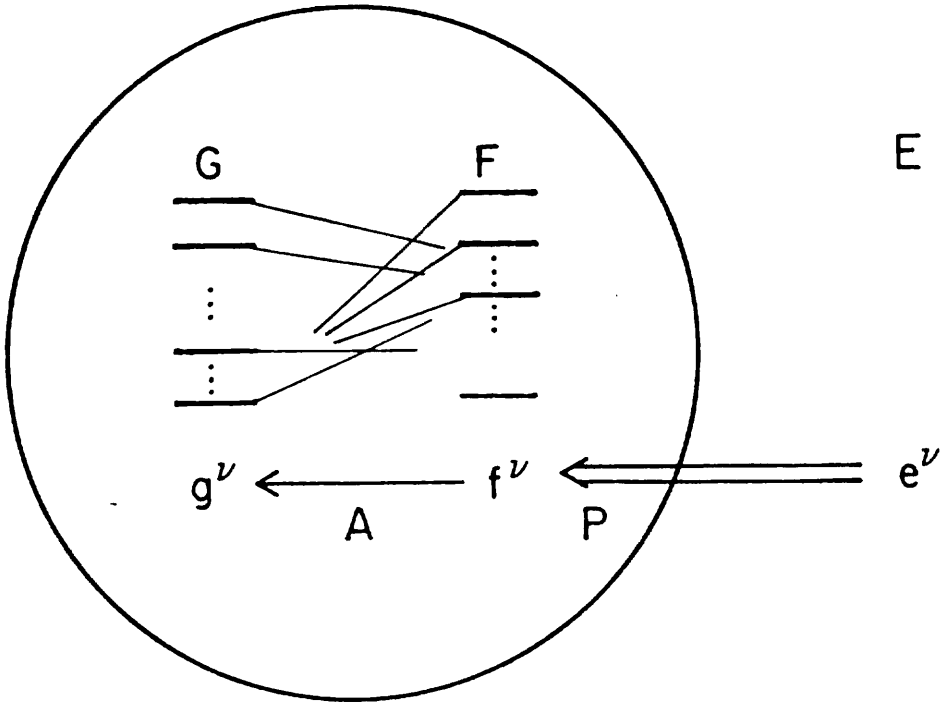


FIGURE 3

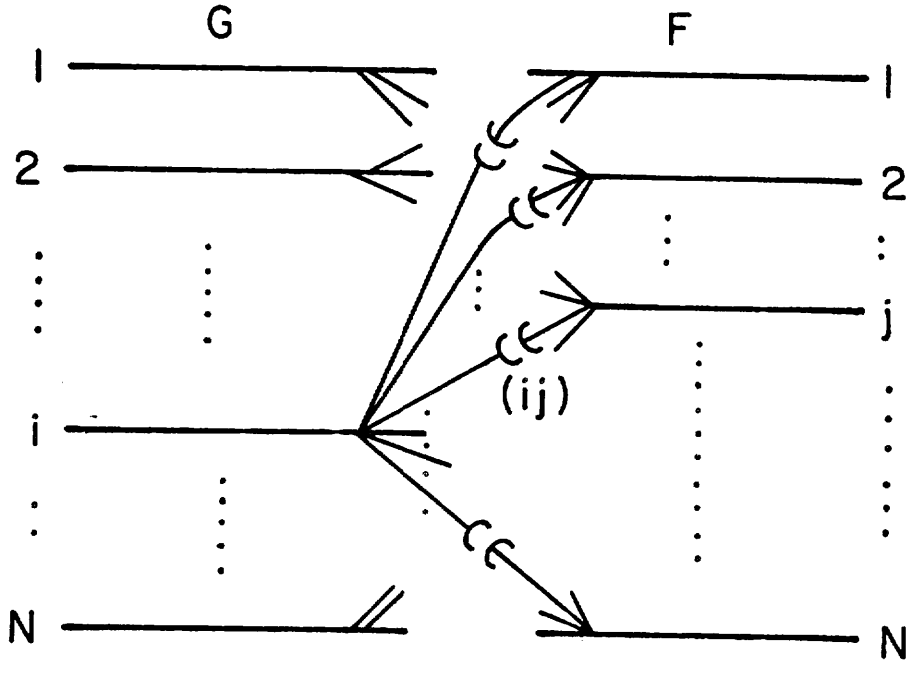


FIGURE 4

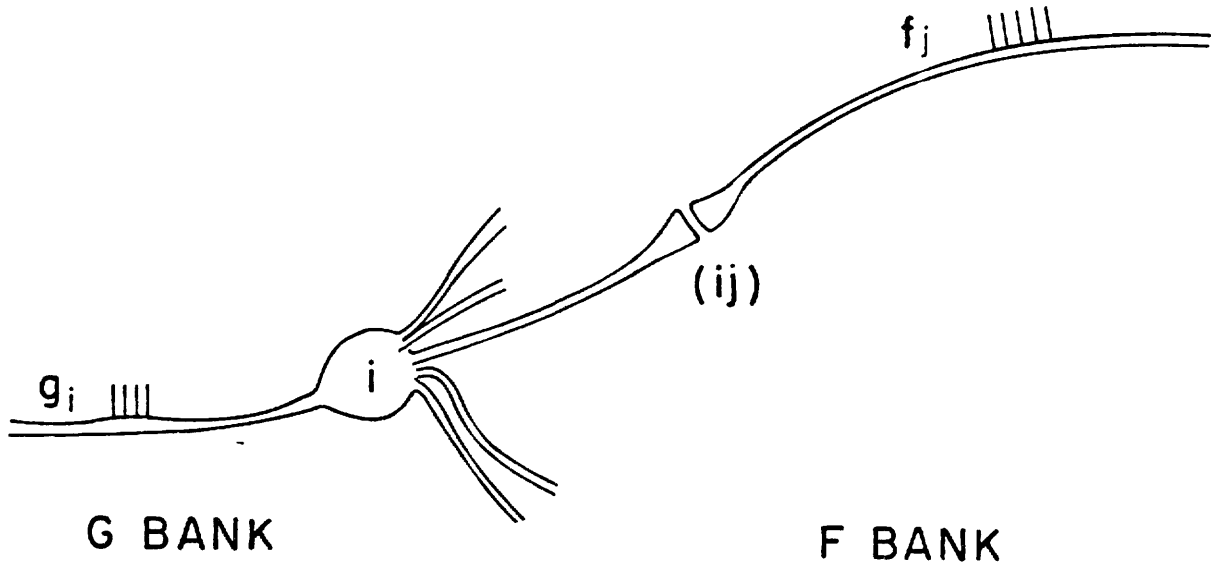


FIGURE 5

FIGURE 6