NEURAL MODELS AND MEMORY[†]

Michael A. Arbib, William L. Kilmer & D. Nico Spinelli

Department of Computer & Information Science
Center for Systems Neuroscience
Technical Report 74C-8
University of Massachusetts at Amherst, MA 01002

[†] Preparation of this paper was supported in part by Public Health Service under Grant No. 5ROI NS09755-03 COM from NINDS and 7ROI MH25329-01 from NIMH.

NEURAL MODELS AND MEMORY[†]

Michael A. Arbib, William L. Kilmer and D. Nico Spinelli
Department of Computer & Information Science
Center for Systems Neuroscience
University of Massachusetts at Amherst, MA 01002

Our basic aim here is to analyze models which bridge the gap between studies of changes in the behavior or overall function of an organism (the psychology of learning) and the study of changes in properties of neurons (as afforded by the neuroanatomy of synaptic change and single-cell neurophysiology). We seek to understand how the brain enables an organism to interact with its world in an adaptive way.

In trying to provide a perspective on neural models and memory, we might have adopted any one of the following strategies:

(i) To provide an exhaustive review of models of neural nets

(ii) To review models of different parts of the brain, and suggest ways in which they might be refined by incorporating memory mechanisms; or

(iii) To provide a "top-down" approach, in which a major effort is placed on what memory models should do, rather than review what current models can do.

Our strategy is a compromise between (i) and (iii). Sections 1 and 3 provide a perspective of kind (iii); while the other sections provide a fragment of the exhaustive review called for under (i). Special attention is given to neural modification in visual cortex; to ethological evidence for learning predispositions; and to models of hippocampal adaptation.

[†] Preparation of this paper was supported in part by Public Health Service under Grant No. 5ROI NS09755-03 COM from NINDS and 7ROI MH25329-01 from NIMH.

1. A PERSPECTIVE[†]

If a neural network is to change its overall behavior, then individual neurons of the network must change, at least in their connections. Thus the problem of training a network can be broken, somewhat crudely, into two parts:

(a) What formulae describe the way in which a neuron (including its connections) will change its behavior over time on the basis of its input-state-output history? [i.e. how may we represent a neuron as an adaptive system?]

(b) Given the adaptive nature of its neurons, how must a network be structured if the cooperative effect of change in the individual neurons is to yield improvement, in some sense, of the function of the overall network?

Of course, these questions reflect but two levels in a hierarchy. The neurochemist will ask what changes in membrane/transmitter/DNA/RNA characteristics will cooperate to yield the neuronal changes of (a); and the neuropsychologist must seek to characterize what constitutes adaptive change in one brain region when it only affects the organism's overt behavior in concert with many other brain regions (as will be suggested by our brief mention of Luria below. We focus on (a) in Section 2 and (b) in Section 3. Then in sections 4 we comingle the two questions as we examine learning schemes posited for hippocampus.

It is inadequate to view an organism as simply responding to a succession of stimuli--rather the internal state of the organism will determine, in great part, to what it will attend (its input) and how it will act upon its

---

[†] For a more general perspective on neural modelling, less attracted to memory problems, see Arbib (1974).

environment (its output). An organism, then, must also model the salient features of its environment and place its activity in the context of dynamic interaction. Below, we shall develop the "slide-box" metaphor for internal models which - with its picture of many slides continually interacting and being updated, with no single locus of exclusive serial processing - serves to emphasize our need to understand how computations may take place in a highly parallel network of dynamically interacting subsystems. Such an understanding may also help us probe the effect of brain damage upon behavior.

In 1929, Karl Lashley published a book "Brain Mechanisms and Intelligence" in which he reported that the impairment in maze-running behavior caused by removing portions of a rat's cortex did not seem to depend on what part of the cortex was removed. He thus formulated two "laws": the "law" of mass action - that damage depended on the amount removed - and the "law" of equipotentiality - that every part of the brain can make the same contribution to problem-solving. Such data have seemed to many irreconcilable with any view of the brain as a precise computing network - but we may effect a reconciliation if we stress (with Luria - see below) the notion of a computation involving the cooperation of many subroutines that are working simultaneously in parallel. Often, a computation can be effected by a subset of the routines. In general, removing subroutines will lower efficiency, though for some tasks the missing subroutines may be irrelevant, so that their removal saves the system from wasting time on them when other tasks are to be done.

Thus equipotentiality is really only valid if we use rather gross measurements of change in behavior - the underlying reality would seem to be the removal of subsystems which can make quite different contributions to a given

level of performance. The brain theorist may thus find intriguing the notion of neuroheuristic programming: structuring a heuristic program in terms of concurrently active subsystems, with anatomical correlates. In this context it is worth recalling Luria's [1973] statement that the concept of localization of function has come to mean a network of complex dynamic structures or combination centers, consisting of mosaics of distant points of the nervous system, united in a common task. Function is understood to be a complex and plastic system performing a particular adaptive task and composed of a highly differentiated group of interchangeable elements. The fact that the elements are interchangeable would allow transfer of function to occur. He goes on to state that the smooth execution of each act or function requires a series of both simultaneously and successively excited connections.

One of Luria's case studies was of a woman who had probably bilateral lesions involving predominantly the parieto-occipital region. She could neither copy letters, nor write them from dictation, but was able to write these letters if they were included in whole well-assimilated words. She could also write the alphabet correctly. In our computer jargon, we might say that it was as if the patient could no longer go from the name of a subroutine to its entry point, but could use the subroutine in those programs which already included the entry point, rather than requiring explicit generation of the entry point anew each time the routine was required. The vividness of this simile encourages our interest in developing models of distributed computation appropriate to brain function.

Within this general context, we may now consider the memory structures required by any system which is to be said to perceive: A human knows that if he presses his hands in a downward direction at a fairly large velocity towards a rectangular surface (a table top) which he has sensed only visually,

his hands will not go through the surface, and a noise will ensue. But it is no trivial task to program a computer to take so little visual information, and make predictions about the trajectory that it could get with its effectors, and the resultant feedback, constraints, and sound. Such predictions are the essence of perception, extrapolating from partial sensory information a great deal in many modalities besides those sensed, that is relevant to action.

Here, then, is the by now well-known idea of a long-term model of the world (see, e.g., Craik [1943], Gregory [1969]), something that tells us that when we see surfaces of a certain kind, they are associated with certain textures, feelings, constraints on action, etc. We further posit that we have a short-term memory which contains information representing our model of the current state of the environment. Rather than every millisecond having to recognize the scene anew, we just notice discrepancies and use this greatly reduced information flow to update the short-term model which guides our activities. It is the long-term model that allows us to build up these short-term models using only partial information to gain access to far more information relevant to action. [This is an analysis-by-synthesis view of perception.]

Our crucial thesis is that perception is inseparable from memory, which in turn is meaningless without reference to the action of the organism - or, more properly, the interaction of the organism with the environment. Perception can be seen as the construction of a partially predictive internal (short-term) model, using long-term memory to incorporate information about action possibilities, and about sensory information from modalities besides those cuing. Perception is dynamic - both in that current information tends to be treated in the context of an existent short-term model, and also in

that the extant model "unfolds" with time.

Clearly, then, information - save at such high levels as involved in human linguistic activity - must be continually referred to what we shall call the ACTION FRAME, namely, the frame of reference induced by the postural and effector mechanisms of the organism, the former serving to stabilize the frame offered by the latter. The repertoire of possible actions of an organism, and of possible questions it may ask, helps to determine the most appropriate neural representation of information. In particular, the organism may encode stimuli in different ways on different occasions depending on the questions prevailing at the time.

Perhaps some more insight into our notions of models of the world may be gained by a metaphor drawn (Arbib, 1972) from the making of movie cartoons. Drawing each frame individually is too inefficient, and so instead use is made of the layering technique. One might go for a whole minute of the cartoon without the background changing, so one could draw it just once. In the middle ground, there might be a tree, say, about which nothing particularly changes during a certain period of time except its position relative to the background. It could thus be drawn on a separate layer, which could then be displaced for succeeding frames. Finally, in the foreground, it may well be that one could draw most portions of the actors for repeated use, and then position the arms, facial expressions, etc., individually for each frame. The layers can then be photographed appropriately positioned in a slide-box for each frame, with only a few parameter changes and minimal redrawing required between each frame.

A similar strategy for obtaining a very economical description of what happens over a long period of time might be used in the brain, with a long-term memory (LTM) corresponding to a "slide file" and short-term

memory (STM) corresponding to a "slide-box". The act of perception might then be compared to using sensory information to retrieve appropriate slides from the file to replace or augment those already in the slide box, experimenting to decide whether a newly retrieved slide fits sensory input "better" than one currently in the slide-box. Also, part of the action of the organism in changing its relationship with the environment might be viewed as designed to obtain input which will help update the STM, by deciding between "competing" slides, as well as helping update the LTM, by "redrawing" and "editing" the slides.

The theory and modelling which occupies us in the rest of the paper focuses on neural adaptation in circumscribed regions of the brain. The fact that we shall close with little feeling for how to "wire-up" a neural "slide-box system" is a measure of how much further both neurophysiology and neural modelling have to go if we are to understand the neural mechanisms of memory. Time and again, we shall be struck by the question of how changes which should be effected at the level of organismic behavior can in fact be yielded by appropriate changes at the neural level. To put it in crude terms, we should continually ask ourselves:

"What is in it for the neuron?"!

Unfortunately, we shall be able to do little more than suggest how synaptic changes might be able to yield adaptive changes in the behavior of a single neural network--the other levels of discussion must be left to other papers. Before turning, however, to the organizational principles which give meaning to the overall behavior within which the activity of such neural networks must be imbedded, it is perhaps worth commenting that vertebrate strategies

of learning by the organism may be essentially different from those of invertebrates. Thus, studies of learning in such invertebrate preparations as Aplysia may only contribute to mammalian studies at the level of determining the mechanisms of neural modification, rather than at the level of seeing how such modifications may contribute to change in network behavior. And, even at the level of component mechanisms, we may run into some trouble since one may caricature the difference between invertebrates and mammals by stressing the uniqueness of cells in Aplysia, and the essentially layered structure of cell tissues in the latter. It may be that visual systems of fly and octopus may provide an appropriate bridge between the two.

One final comment about the problem of level, as we try to relate the overall function of an organism to changes in the individual neurons; that is the gross difference in time scale, where a human may learn complex intellectual structures on a time scale from seconds to years, whereas the individual action of neurons is on a millisecond time scale. We shall get some feel for this when we look at the convergence schemes for perceptrons in the next Section--but it must be stressed that a full understanding of human learning cannot proceed without ideas whose faint glimmerings are seen in the data structures being evolved by workers in artificial intelligence, as in such works as Schank & Colby (1973).

2. TWO BASIC SCHEMES FOR NEURAL LEARNING.

In this Section, we shall study two mathematical models of neural change. To focus the discussion, we consider the system shown in Fig. 1. We view the preprocessor as a mechanism that extracts from the environmental input a set of d real numbers. The set will be called a pattern and the numbers components of the pattern. The pattern recogniser then takes the pattern and produces a response which may have one of N distinct values where there are N categories into which the patterns must be sorted.

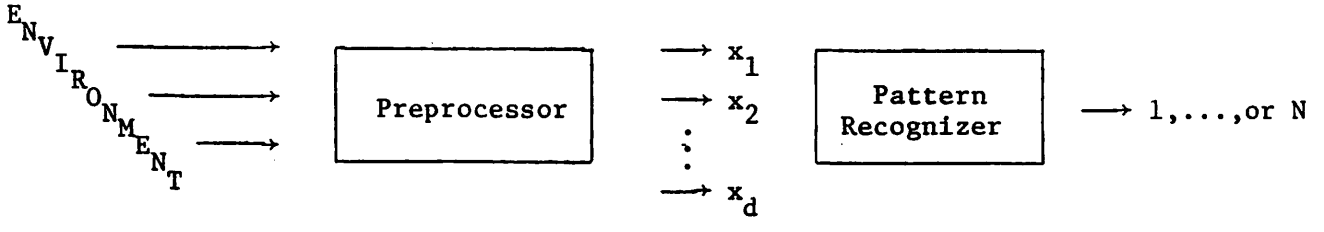

Figure 1

The two classic learning schemes for McCulloch-Pitts type formal neurons are the Hebb [1949] scheme (strengthen an active synapse if the efferent neuron fires) and the Perceptron scheme of Rosenblatt [1961] (strengthen an active synapse if the efferent neuron fails when it should have fired; weaken an active synapse if the efferent neuron fires when it should not have fired).

The Hebb scheme has been elaborated by Brindley [1969] whose ideas have been developed by Marr [1969, 1970, 1971] in models of cerebellum, hippocampus (see Section 4) and neocortex which provide ingenious mechanisms of codon selection, adjustment for level of background activity, etc. While these constitute an important contribution to our growing array of tools

for the synthesis of neural networks with specified properties, we remain sceptical of their value as an analysis of the given regions of brain. Grossberg (see, e.g., [1972]) has many studies of learning networks, with components ranging from the neural to the psychological, which make rich contact with studies of conditioning and of motivation. Kilmer and Olinski's [1974] study of learning in hippocampus (Section 4) is especially interesting for its use of developmental stages in tuning up the system.

The Hebb model has also been used by von der Malsburg [1974] (see also Perez, Glass and Schlear [in press]) in his model of the development of line detectors in cat visual cortex. He uses a normalization rule; and a lateral inhibition to stop the first "experience" from "taking over" all "learning circuits" which was used earlier by Spinelli [1970] in his OCCAM model of spinal-cord like circuitry for an associative memory. In fact Hirsch and Spinelli [1970] have shown that cat visual cortex is not restricted to line detectors, but can take the imprint of a specific visual experience, at least if the kitten sees only one or two things. We shall consider the work of Spinelli and von der Malsburg in Section 2.1.

The Perceptron model was put on a firm mathematical footing by Block [1962]--and has been embedded in a good textbook treatment of "Learning Machines" by Nilsson [1965]. We shall study the Perceptron convergence theorem in Subsection 2.3. Minsky and Papert [1969] have shifted attention from convergence questions to "what can a given network learn?"; questions which have an interesting relationship with the network complexity studies of S. Winograd [1965] and Spira [1969]. We shall place their work in the context of network predisposition in Section 3.2.

The hologram as a mechanism for associative memory, and the whole range

of Fourier optics, has provided much stimulus for neural modellers, although we shall not develop it in detail here. Gabor et. al. [1971], Willshaw Buneman and Longuet-Higgins [1969], Kohonen [1971] and Westlake [1970] are among the many who have molded features of holography into neural networks. However, we think that those who, like Pribram [1974], or Pollen and Taylor [1974], expect the whole gamut of mathematical techniques--"the visual system makes a Fourier transform"--to carry over from optics to neural nets are mistaken. Such a view seems ill suited to handle our ability to perceive the world as made up of independently moving objects. Again, the reconstructibility of the hologram differs from our act of perception as a preparation to act, as distinct from storing a veridical image. Blakemore et. al. [1970], Campbell et. al. [1968] and others have stressed the role of spatial frequency in the visual system, but it seems to us that these can provide cues (e.g. for texture) which augment cues such as contour, rather than providing a total frequency transform prior to "pattern recognition". However, we will not develop the holography theme further in this article.

2.1 Learning without a Teacher.

In this section we consider Hebb-type schemes in which synaptic weights are adjusted without explicit reinforcement.

2.1.1 Experiments on Visual Memory†

It has been recently discovered that by limiting the visual experience of kittens to one or two simple visual patterns, it is possible to fill up, as it were, their visual cortex with receptive fields whose shape resemble the patterns in one or more details; some receptive fields are even recognizable, though blurred, representations of the image seen by the cat. Atrophy from dis- or misuse of visual mechanisms is well known in the clinic. Moreover, the experiments of Hubel and Wiesel (1965, 1970) had shown that monocular suturing or squint causes loss of binocularity of cells in the visual cortex of cats whereas binocular suturing did not. These findings combined to suggest that by having kittens view a pattern of (three, say) vertical lines through one eye and a pattern of horizontal lines through the other during the critical period of development, one might be able to cause loss of binocularity for cells with vertical and horizontally oriented receptive fields as they were fed discordant information. To rephrase: the vertically exposed eye was expected to have a "hole" in the distribution of orientations for verticals. Appropriate "holes" in the behavioral repertoire, testing one eye at a time, would have betrayed the function of the missing units.

The effects that this simple procedure had on the functional properties of visual cortex cells (Hirsch & Spinelli, [1970, 1971]); were of astonishing

† For further information on this topic, see the papers by Pettigrew and by Spinelli elsewhere in this volume.

magnitude. There were only two types of units in the visual cortex of a kitten so deprived:

1) Units uncommitted to specific visual features. These units did not have a receptive field mappable by Spinelli's automated method and did not show any selectivity to lines and edges under manual control;

2) Units with elongated receptive fields. There were only two orientations. Units with vertical orientation could be mapped and/or would respond to vertical lines only through the eye that had seen the three vertical bars; units with horizontal orientation could be mapped and/or would respond to horizontal lines only through the eye that had seen the three horizontal bars.

Even more astounding were a few units with receptive fields that looked like a carbon copy of the stimuli viewed during development! This almost photographic reproduction of images seen by the kittens weeks before suggested that one might be dealing with memory traces. Three possibilitites had to be examined:

1) The various classes of receptive fields one finds in the adult are genetically preprogrammed. Presence at birth, maturation after birth, or the necessity of environmental stimulation for the genome to express itself are all subsumed in this hypothesis and could in various degrees appear in the same species for various classes of units or in different species. The hypothesis can then be made that the unstimulated units atrophy, fail to mature or are not expressed. This hypothesis demands clear and predictable behavioral deficits from the kittens described above. It also predicts that once the

damage is done, it should be permanent, i.e. letting the kittens have further, normal experiences after the critical period should not change the physiological picture. Further, the set of available classes can be reduced but not changed.

2) Cells are genetically preprogrammed as in (1). However, during the critical period, partial or total reprogramming under environmental control can take place if needed; this would insure that the animal has feature detectors optimal for the environment it finds itself in. In the worst case, it would have a set of detectors, that which has proven most helpful to its species through natural selection. The transience of the critical period would be an advantage since it would be impossible for the rest of the brain to interpret information from detectors whose coding properties change over time (cf. Kilmer's core vs. non-core scheme for training hippocampal circuitry, in Section 4). This hypothesis has essentially the same predictions as the one above, i.e. clear behavioral deficits for tasks that demand nonexisting detectors and permanence of the physiological effects; however, it differs from the above in that even though there would be limits, it allows for the property of generating receptive fields totally different from the ones originally preprogrammed.

3) This could be called a memory hypothesis, i.e. there are no genetically programmed detectors: what is programmed is an adaptive network which is capable of storing, in a very direct fashion, elementary visual experiences. The receptive field shapes one maps in an adult cat would be bits and pieces of what the animal has seen in his past. This hypothesis predicts changes in the physiological picture, if

new experiences are allowed, and also predicts no or slight (learning capacity does decrease with age) behavioral deficits.

In real life, of course, these three possibilities would be present in various ratios depending on the animal.

Shortly after the Spinelli-Hirsch experiment appeared in the literature, Blakemore and Cooper (1970) published a similar experiment, though with notable differences. They raised two kittens in the dark except for a few hours every day when one kitten was put in a cylinder painted with vertical stripes and the other in one with horizontal stripes. Recording from the visual cortex after development showed that cells responded best to vertical lines, or close to it, in the vertically exposed kitten and to horizontal lines, or close to it, in the horizontally exposed kitten. Units were binocularly activated.

It is Spinelli's working hypothesis at this time that the experience-shaped receptive field map represents nothing less than that engram after which Lashley searched in vain so long ago: to prove or disprove this hypothesis vis a vis other possible interpretations, and to understand the adaptive structure that brings this about, is the task for his ongoing experiments. We now turn to models related to this phenomenon.

2.1.2 Models of Visual Memory.

Spinelli's [1970] OCCAM (a computer model for a Content Addressable Memory in the central nervous system) suggests a scheme for how memories might be multiply stored in discrete neural networks. The basic neural circuit is reminiscent of spinal mechanisms: there a specific, prewired input pattern, e.g., an itch, elicits a specific, prewired behavior: a scratch reflex. In OCCAM things are similarly organized except that both the input and the output pattern are arbitrary and determined by experience. Further the model specifies how inputs "find their way" to the appropriate memory trace, i.e., the OCCAM networks are content addressable. Very little, if any, preprocessing of sensory activity is assumed. The model posits that the simplest and safest thing for an organism is to "store" information as it arrives, "selection" of "meaningful" stimuli is done by biasing memory at the time of action. Anything can thus become important and facilitate its own subset of memories.

Consider, then, the memory networks of Figure 2, addressed in parallel by stimuli entering the CNS. We shall see how they may form a content-addressable memory, wherein providing the system with part of a chunk of information will enable the system to play back the whole chunk.

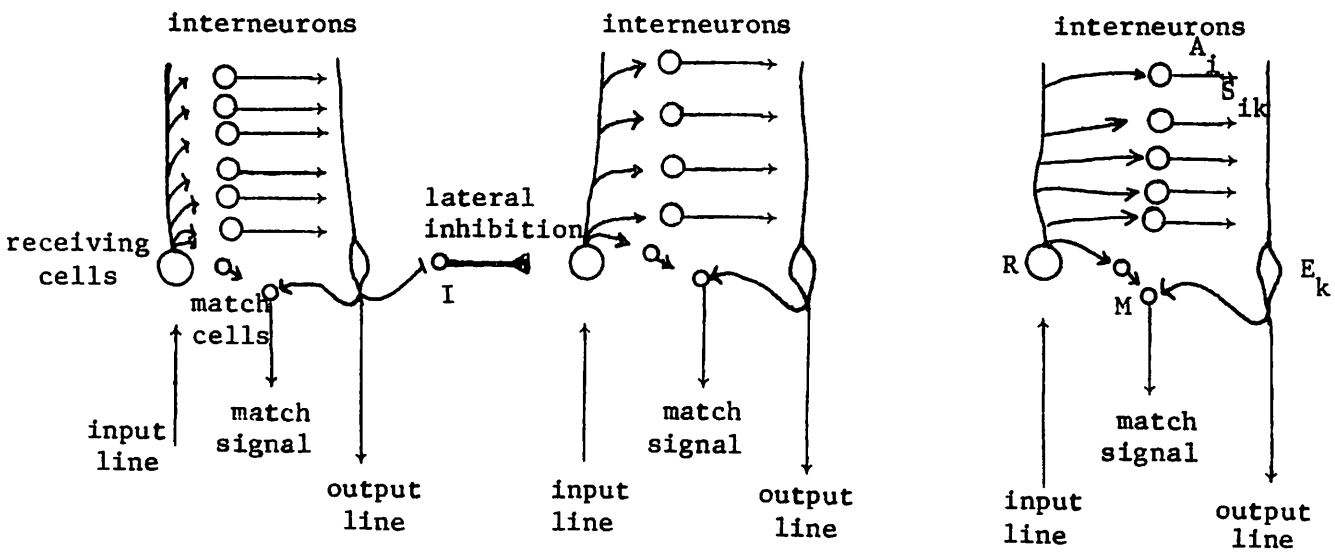

Figure 2.

The different columns are connected by collaterals from receiving cells and match cells which carry lateral inhibition to other input cells in nearby networks. It is assumed that only one interneuron per column is active at any time, and that different temporal segments of a pattern will be switched in a regular fashion through different interneurons. [In motor nerves, where individual fibers fire at about 10 cps, smooth contractions are obtained by regular phasing in and out of motor units.] A crucial assumption is that the synaptic conductivity tends in the limit to be directly proportional to the activity which is going through the synaptic junction itself, so that if a given quantity of activity is presented to the same synapse over and over again, an asymptote will be reached such that the conductivity will represent faithfully the amount of activity that produced it.

We then assume that <u>whenever a synaptic connection is activated, the amount of excitatory potential generated is proportional to the synaptic conductivity</u> (not to the activity that generates it). If a temporal pattern is presented to a network repeatedly, it will be stored in the synaptic conductivity of the interneurons, which will thus cause the output cell to play out a better simulacrum of the input pattern.

The match cell output provides a measure of the correlation in the activity of the input cell and output cell from which it receives collaterals. The pattern is presented to all networks in parallel. But eventually, one cell will have a somewhat better adjustment than its neighbors, the activity in its match cell will rise and (thanks to the lateral inhibition mechanism) turn down the input to nearby networks--the network that gets ahead, by chance, draws the pattern to itself, and prevents the other networks from learning it. The number of networks that learn the same pattern is thus determined

by the extent of the lateral inhibition. To ensure that the lateral inhibition "gets there in time", it is assumed that each time a cell is activated, an afterdischarge occurs--and that the longer the afterdischarge, the more the synaptic conductivity will be changed. Thus, the longer the afterdischarge, the faster the learning--and this afterdischarge will be cut off by lateral inhibition if another "column" has already learnt the pattern.

When one or more patterns have been stored, it is desirable that if new patterns are to be stored, this be done by networks that have not been previously used. To do this a given match cell becomes harder to activate if its network has been used often before. Then a pattern would have above-chance effect on a network in which it was already stored, reasonable effect upon an 'uncommitted' network, and little effect on a network containing a well-learned pattern.

Since a portion of a pattern correlates well with the whole output pattern, the match signal can signal which output cell is playing back the pattern of which the input is a fragment. A stimulus might then lead to the playing back of a sequence that includes both receiving the stimulus and making the adequate response. It should be pointed out that memory will also contain sequences where the organism performed a response that led to undesirable consequences, however, here is where the power of content addressable memories comes in: a bias for the desired event will facilitate sequences that contain it. Acceptable match indicates what portion of the input pattern must match the output pattern for the correlation to be significant.

Pribram, Spinelli and Kamback [1967] suggest that presentation of a stimulus will generate a playback of the whole sequence: recognition of the stimulus, the appropriate behavior that went with the stimulus, followed

by the expectation of the consequences of that behavior. The less of the stimulus that is presented, the more information is in the playback, and the more the risk is in using it. [But an animal might generate actions which experiment to see if the situation accords with the recalled details of the stimulus before reacting to the stimulus per se.] Ideally, then, the acceptable match parameter should be set for that minimum value which allows unequivocal recognition of the stimulus and thus the playback of the rest of that memory package containing information about what to do or not to do with it and what to expect.

The model assumes that while visual memory contains primarily visual information, it also contains enough non-visual information to allow the read-dressing of the system by the visually triggered memories so that auditory, somatic, gustatory, etc., strings are subsequently called into play. The internal addressing of the memory by internal states, which would be part of the string, for example hunger and the disappearance of hunger, would activate or would facilitate all those memory strings that contain such information in themselves, and therefore produce a partial level of match. This would then make available to the rest of the brain strings containing pertinent information about feeding behavior.

The following is the <u>key</u> to how such a memory structure may serve the adaptive behavior of the organism:

> If other parts of some strings are available in the environment, a higher level of match would be achieved for certain strings and the connected behavior could then be played back if the acceptable level of match is reached or exceeded. Memory is thus continuously addressed by three agencies: internal states, external stimuli, and recently activated memories; it is the interplay of these three factors that give behavior its continuity, variety and purposefulness.

How would OCCAM [Omnium-gatherum Core Content-Addressable Memory] store patterns which are very similar, but have different meanings, as different patterns? "Key" endings to the two patterns would of course make them different. Perhaps this is the way in which reinforcers act--behaviorally and neurally two patterns which might have looked identical are really different, because the consequences, which are part of the same memory package, are different.

Reinforcers could also act to decrease lateral inhibition and increase learning speed, so that an organism might learn faster and more redundantly those strings whose information had survival value. Such reinforcers as pain and food might be permanently wired-in, others would act on memory only through the software, as parts of existing programs or Plans--cf. Pribram (1969).

Whereas the OCCAM model of 1970 had a temporal-spatial converter to enable it to record temporal patterns, the von der Malsburg model [1973] is closer to the spirit of the results on visual pattern memory in Section 2.1.1, being formulated to determine whether a simple circuit possessing only a few characteristics of the cat's visual system would organize itself into the "simple cell" receptive field patterns found by Hubel and Wiesel in area 17 of cat visual cortex, each cell having one preferred orientation to which it responds maximally. Malsburg thought that genetically specified patterns of lateral excitatory and inhibitory influences in the geniculo-striate system highly predispose this system toward its columnar organization. He also thought that a loose genetic specification of the details of the retino-geniculo-striate projection could be coupled with plastic synaptic mechanisms in the projection so as to enhance the columnar organizational

tendencies of the striate system.

His model to test these ideas consisted of a retina of 19 binary elements, a geniculo-striate manifold of 169 excitatory cells (E cells) and 169 inhibitory cells (I cells) interconnected according to a simple geometric specification. (The inhibitory interconnections serve much the same role as the lateral inhibition in OCCAM, with the E-cells being the "match cells" and "output cells" since 'recognition' rather than 'temporal playback' is the task here.) The connection from each retinal cell A_i to each E cell E_k is through a Hebb synapse of strength S_{ik} which is modified by $\Delta S_{ik} > 0$ each "modification time step" if A_i is active and if E_k fires. In this case the magnitude of ΔS_{ik} is proportional to the firing rate of E_k.

The firing rate of any E or I cell is equal to the amount by which the excitatory state H_k of the cell exceeds its threshold θ_k. Defining H_k^* as H_k if $(H_k - \theta_k) > 0$ and as 0 otherwise, the equation for H_k is

$$\frac{d}{dt}H_k(t) = -\alpha_k H_k(t) + \sum_{i=1}^{338} W_{ik} H_1^*(t) + \sum_{i=1}^{19} S_{ik} X_i(t),$$

where: α_k = decay constant of H_k.

W_{ik} = synaptic strength from E or I cell i to cell k.

$X_i(t) = 1$ if A_i is active and 0 otherwise.

Since all $\Delta S_{ik} > 0$ are positive, Malsburg had to renormalize his circuit to return $\sum_{i,k} S_{ik}$ to C(C a constant) after each "modification time step" in order to keep his H_k bounded. He presented each retinal input with his model in a relaxed initial state, and then waited until all H_k reached

equilibrium (which always happened within about 20 time steps) before he designated the next time step as a "modification time step."

He used nine different retinal inputs, each corresponding to a "line" of different orientation. He presented them in an appropriately mixed order, and after 100 modification times froze the ΔS_{ik} and checked the retinal input orientations that elicited maximal responses in each E cell. Fig. 3 shows the result, which is strikingly reminiscent of Hubel and Wiesel's columnar mapping results. Malsburg's model actually did organize itself, with the help of a very restricted set of retinal inputs, because initially the preferred orientation map was chaotic and not at all like Fig. 3. Though Malsburg got clusters of high E cell activity from his intrinsic geniculo-striate dynamics before any S_{ik} adjustments were made, his high orientation specificity was accomplished by S_{ik} learning.

Malsburg's model has rather severe limitations that are understood at this stage only qualitatively. His scheme is based on a highly restricted set of retinal inputs. His E-to-I and I-to-E influences are additive, non-plastic, and spatially arranged according to simple geometric rules. In reality, the brain's ubiquitous lateral connections must often convey signals that exert subtle logical effects, for example in frontal granular cortex of primates, where signals from all over the rest of the brain are brought together. Also Malsburg does not use any evaluative criteria for changing his synaptic strengths other than Hebb's rule that tends only to further entrench learning that has already occurred. His model could never reverse an already well learned response pattern for example.

Both the OCCAM and Malsburg schemes provide local, but multiply represented, storage. Both will store <u>whatever</u> comes in, not just the lines that Malsburg used in his test (cf. Lenherr [1974]). Lateral inhibition, which was not

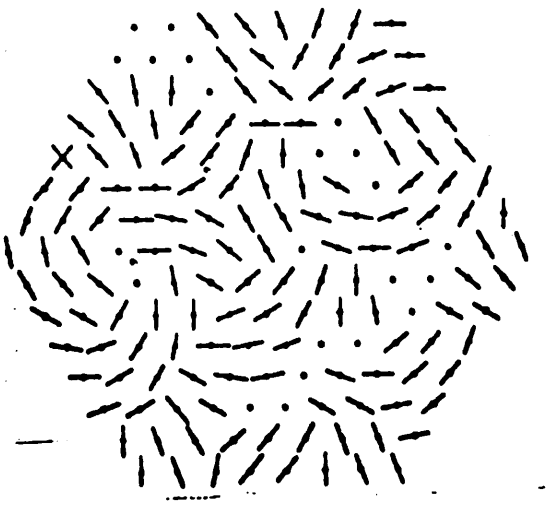

Fig. 3. Each bar of this view onto the cortex indicates the optimal orientation of the E-cell. Dots without a bar are cells which never reacted to the standard set of stimuli. Two bars indicate two separate sensitive regions. (Malsburg's Fig. 13.)

mentioned in the basic experiments of Hubel and Wiesel, plays a crucial role in the models, so that different "experiences" will be stored in different units. In such systems, feature detectors arise only to the extent that they respond to commonly occurring aspects of the animal's experience.

As new experimental findings accummulate they will tell us how to merge the various models into one which will account for what is known. In the meantime models are a continuing source of inspiration for new experiments and experiments are a continuing source of inspiration on how to improve models. This process of successive approximations is bringing us closer and closer to understanding how real brains work. Sharp theories and good experiments combined with simulation to provide testability, now that we can trace the effects of experience, will certainly unravel these adaptive neural networks that evolution took so long to develop.

2.2. The Perceptron.

In the terms of Figure 1 of this section, a pattern recogniser is a function $f: \mathbb{R}^d \to \{1,\ldots,N\}$. The points in $\mathbb{R}^d$ are thus grouped into at least N point sets which we shall assume can be separated from each other by surfaces called decision surfaces. We shall assume for almost all points in $\mathbb{R}^d$ that a slight motion of the point does not change the category of the point. This is a valid assumption for most physical problems. The additional problem still exists of the category that is represented in more than one region of $\mathbb{R}^d$. For example, a, A, α,$\mathring{A}$, are all members of the category of the first letter of the English alphabet, but they would probably be found in different regions of a pattern space. In such cases it may prove to be necessary to establish a hierarchical system involving a computer apparatus that recognises the subsets and a separate system that recognises that the subsets all belong to the same set. At any rate, let us avoid this problem by assuming that the decision space is divided into exactly N regions, eliminating split categories.

We call a function $g: \mathbb{R}^d \to \mathbb{R}$ a discriminant function if the equation $g(x) = 0$ gives the decision surface separating two regions of a pattern space. A basic problem of pattern recognition is thus to specify such functions. Unfortunately it is virtually impossible for a human to "read out" the function he uses (and in what way?!) to classify patterns. What, for example, is your intuitive idea of the appropriate surface to discriminate A's from B's? So a common strategy in pattern recognition is to provide a classification machine with an adjustable function, and "train" it with a set of patterns of known classification that are typical of those that the machine must ultimately classify. The function may be linear, quadratic or polynomial depending on the complexity and shape of the pattern space and necessary

discriminations. Actually the experimenter is choosing a class of functions with adjustable parameters, which he hopes with proper adjustment will yield a function that will successfully classify any given pattern. For example, the experimenter may decide to use a linear function of the form:

$$g(x) = w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_dx_d + w_{d+1}$$

in a two-category pattern classifier. The equation $g(x) = 0$ gives the decision surface, and thus training involves adjusting the coefficients $(w_1,w_2,\ldots,w_d,w_{d+1})$ so that the decision surface produces an acceptable separation of the two classes. We say that two categories are <u>linearly separable</u> if in fact an acceptable setting of such linear weights exists.

The reader may regard adaptive training as a case of the identification problem - it is as if we were trying to find a model of a black box which classifies the patterns on the basis of some samples of its input-output behavior.

Consider the case of a two-fold classification effected by using a threshold logic unit (= McCulloch-Pitts neuron) to process the output of a set of binary feature detectors. We then have a set R of input lines (to be thought of as arranged in a rectangular "retina" on which patterns may be projected) for a network which consists of a single layer of neurons whose outputs feed into a threshold logic unit with adjustable weights. We want to analyse what classifications of input patterns can be realised by the firing or nonfiring of the output of such an array given different weight settings.

Such a net is an example of what Rosenblatt [1961] calls a Perceptron, and, as we have said, is used to classify patterns on the retina into those which yield an output 1 and those which yield an output 0. The question asked by Rosenblatt and answered by many others since (an excellent review is in Nils Nilsson's [1965] monograph on 'Learning Machines') is:

"Given a network, can we 'train' it to recognise a given set of patterns by using feedback, on whether or not the network classifies a pattern correctly, to adjust the 'weights' on various interconnections?"

The answers have mostly been of the type:

"If a setting exists which will give you your desired classification, I guarantee that my scheme will eventually yield a satisfactory setting of the weights".

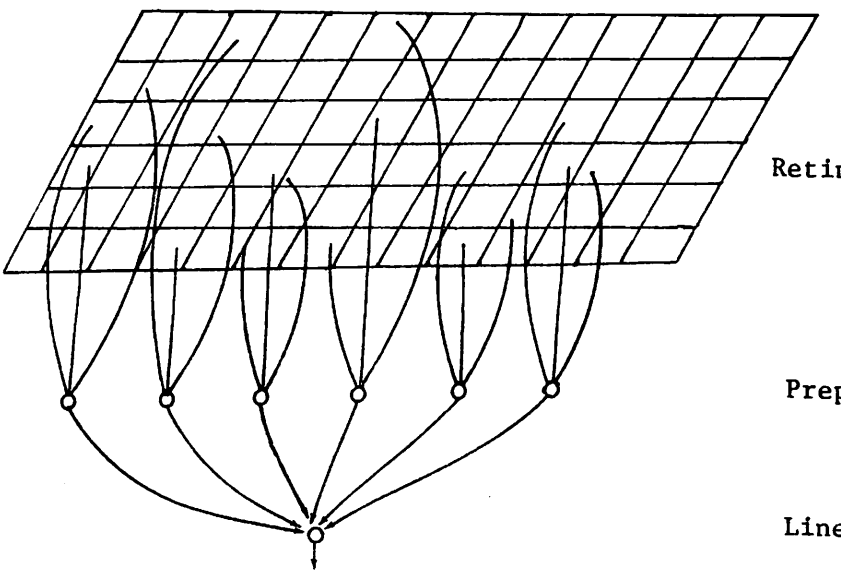

Figure 2

We now give one of the Perceptron convergence schemes. First some notation: With each predicate ψ we shall associate the binary function

$$\ulcorner\psi(x)\urcorner = \begin{cases} 1 & \text{if } \psi(x) \text{ is true} \\ 0 & \text{if } \psi(x) \text{ if false} \end{cases}$$

We recall, too, that for two vectors $x = (x_1,\ldots,x_n)$ and $w = (w_1,\ldots,w_n)$, we use $w\cdot x$ for the scalar product $(x_1w_1 + \ldots + x_nw_n)$.

Suppose, then, that there are d feature detectors in the preprocessing layer, so that the input to the linear threshold unit is a vector $x = (x_1,\ldots x_d)$. Let us augment x by adding a $(d+1)^{st}$ component set to 1 to obtain $y = (x_1,\ldots,x_d,1)$. Then if we let $w = (w_1,\ldots,w_d,-\theta)$ which is the weight vector augmented by minus the threshold, we see that our equation for the response r of the unit can be abbreviated from

$$r = 1 \quad \text{iff} \quad \sum_{i=1}^{d} w_i x_i \geq \theta \quad \text{iff} \quad \sum_{i=1}^{d} w_i x_i - 1\cdot\theta \geq 0$$

to the simple form

$$r = \ulcorner w\cdot y \geq 0\urcorner$$

Let us be given a finite set y_1 of augmented vectors corresponding to category 1, and a finite set y_2 of augmented vectors corresponding to category 2. In assuming that the two categories are linearly separable, we guarantee that there exists at least one $(d+1)$ weight vector $\hat{w}$ such that

$$\begin{cases} \hat{w}\cdot y \geq 0 & \text{if } y \in y_1 \\ \hat{w}\cdot y < 0 & \text{if } y \in y_2 \end{cases}$$

We start with an arbitrary weight vector w which presumably misclassifies many patterns, and try to adjust it by repeated application of some

error-correction training procedure. The procedure we shall study works repeatedly through the patterns in y_1 and y_2, testing each to see if the latest w classifies it correctly. If w classifies the current y correctly, we leave w unchanged and move on to the next y. If the classification is incorrect, however, we change w to w' where

$$(*) \quad \begin{cases} w' = w + y & \text{if } y \text{ belonged to category 1} \\ w' = w - y & \text{if } y \text{ belonged to category 2} \end{cases}$$

The idea is as follows:

If y is in category 1 but w misclassified it, then we had $w \cdot y < 0$ where we should have had $w \cdot y \geq 0$. Since $y \cdot y > 0$ for any non-zero vector, we have that $(w + y) \cdot y = w \cdot y + y \cdot y > w \cdot y$ and so - even if we do not have $w' \cdot y \geq 0$ - we at least have that w' classifies y "more nearly correctly" than w does. Similarly for the category 2 correction.

Unfortunately, in classifying y "more correctly" we run the risk of classifying another pattern "less correctly." However, it can be proved (see, e.g., Nilsson [1965]) that our procedure does <u>not</u> yield an endless seesaw, but will eventually converge to a correct set of weights <u>if one exists</u>.

We close this section by noting that Minsky and Papert [1969] have revivified the study of Perceptrons by responding to such convergence schemes with the basic question: "Your scheme works when a weighting scheme exists - but when does there exist such a setting of the weights?" In other words, they ask: "Given a pattern recognition problem how much of the retina must each associator unit 'see' if the network is to do its job?" They analyse this question both for "order-limited Perceptrons" in which the "how much" is the "number of input lines per component"; and "diameter-limited Perceptrons"

in which the "how much" is the diameter of the input array from which each component receives its inputs. We shall say more about this in Section 3.2

3. WHY RATS CAN'T LEARN BIRDSONG

In Section 3.1, a consideration of ethology shows how the neural networks of different animals are predisposed to handle different ranges of problems (rats can't learn birdsong), and to be differentially "tuneable". Then in Section 3.2 we look at the pretheory of network predispositions, complexity, and tuneability. A crucial question for neural modelling must be the study of complexity of computation - for instance, to understand how long a network of given components must take to compute a certain function, or what the range of functions is that can be computed by networks of a given structure. The mathematical theory cannot at present be "plugged in" to solve biological problems, but it may help us refine the questions we ask of the experimenter, and suggest important new ways of interpreting his results.

3.1. Ethology of Network Predisposition.

The relative role of genetic, predispositional, and learning factors in animal behavior is perhaps most clearly suggested by reviewing some results (cf. Marler [1973], Tinbergen [1972], and Ploog [1971]) on the ontogeny of birdsong.

(1) Some birds (e.g., domestic fowl) will produce the species-specific call even if deafened at birth prior to exposure to the call. This implies a pre-wired, genetically determined motor model for a call production. There is no process of matching to an auditory input.

(2) Some birds (e.g., song sparrows) can develop the normal song even though raised in isolation from it, but, if deafened at birth, the normal song fails to develop, suggesting that auditory feedback from their own song production is necessary, presumably for matching with a template.

(3) In the third type of ontogeny, not only is auditory feedback necessary, but also an external 'model' of the correct song must be presented to the developing bird. For example, chaffinches raised in isolation have abnormal songs which lack much of the structure which forms when they are exposed to the species-specific song during development.

(4) Some birds are more open to environmental and sociological influences. An example is the young male bullfinch, which by virtue of the proximity of its father, acquires it species specific song, but if raised experimentally by a foster-species, will instead develop the foster species' song. Such mimetic learning is seen in many parasitic birds which acquire the song of the host early in life, while other parasitic species, such as the European cuckoo, in contrast remain "resistant" as fledglings and retain their own species song. Thus, a complex interaction between an innate susceptibility

(which provides the basis for the elaboration of a song template and which in turn facilitates the development of a species-typical song pattern), environmental influences in the form of exposure, learning, feedback stimulation (reafferentation) while singing, and acculturation or acquisition of a "local dialect" can occur in some species while other avian species show a lesser dependence upon such interacting ontogenetic influences.

We see from the above that some animals are genetically predisposed to learn certain things at certain times during their lives. Spinelli's work - reviewed in Section 2.1.2 - also shows that early learning may occur in nerve cells without any obvious reinforcement. The ethology of imprinting shows that some kinds of probably permanent learning occur within narrow time limits during the lives of individuals. This is often illustrated by noting that a young animal must establish a strong bond to its mother and be willing to put forth enormous efforts to follow her if it is to continue to receive the food, protection, and shelter she can provide.

We shall next give examples in support of the idea that each animal is genetically prepared to learn some things easily and other things, of about equal complexity by our standards, not at all.

Garcia and Robert A. Koelling (Garcia [1971]) confronted rats with a sweet tasting liquid and a light-sound stimulus, both paired with radiation sickness--a malaise characterized by stomach upset. But only the taste, not the bright light or loud sound, became unpleasant to the rats. In the complementary experiment, they paired the sweet taste and the light-sound stimuli with electric shock. This time, the rats associated the light-sound combination, not the taste, with the shock.

Rat pups behave similarly: they avoid nursing from a foster mother if

made sick from an injection shortly afterwards, whereas if not made sick they will nurse freely and repetitively from their foster mother.

The above adult rat experiment illustrates both ends of the learning predispositional spectrum. The rats were genetically predisposed to associate taste with illness, and this association occurred in spite of the hour-long delay between the taste and the illness. But the rats were not predisposed--were perhaps negatively predisposed--to associate taste with electric shock and to link external events (light and noise) with nausea. The evolutionary advantage is obvious: animals that are poisoned by a distinctively flavored food and survive, do well not to eat that food again.

We next quote from the article by J. Garcia [1971]..

> "The remarkable ability of thiamine-deficient rat to use its internal feelings as a detection device to search for food containing thiamine was demonstrated by Paul Rozin and his associates at the University of Pennsylvania. When the symptoms of thiamine deficiency become evident, the rat will change its eating pattern. It will tend to give up its current diet and to explore new food sources. When placed in an 'experimental cafeteria' containing a variety of flavoured foods, one of which contains thiamine, the rat will go into a 'testing mode'--that is, it will eat small meals of one new food at a time, and space its meals several hours apart. In this way it soon locates, and then concentrates upon, the food containing thiamine. Dr. Rozin says that the rat acts precisely as would a rational man who had lost all the labels in his medicine cabinet and was feeling ill.
>
> . . .
>
> "Bennett Galef of McMaster University, Ontario, has shown that the feral rat is also much more responsive to novel stimuli and to bait-shy tests than the laboratory rat. Using his experimental enclosures, he has demonstrated that when wild rats are made bait-shy to a specific food source and allowed to raise a litter of young, these offspring also tend to avoid that food source even when they have not had the flavour-illness experience. Somehow the parents are able to communicate this information to their young [cultural transmission!].
>
> . . .

> A wide variety of conditioning studies have given rise to the principle of immediate reinforcement, which implies that rewards and punishments should be applied immediately in order to be effective. The ability of an animal to associate consumption of a meal with its effect hours later is a notable exception. It has been postulated that this apparently unique ability is due to the peculiar nature of food stimuli. Traces of the meal or drink linger in the mouth and gastrointestinal track for hours and are there insulated from interfering stimuli in the post-feeding period; thus, they could physically bridge the gap between ingestion and illness.
>
> Three kinds of evidence argue against this hypothesis. First, Samuel Revusky of Northern Illinois University, DeKalb, has shown that the rat will not necessarily reject the last food item consumed but rather it rejects the last novel item consumed. Moreover, it will not ordinarily develop an aversion for familiar foods intervening between the consumption of a novel food and the subsequent illness. In another study, Dr. Revusky gave his animals sucrose in water and made them ill more than seven hours later. Presumably, all traces of sucrose should have vanished from the gastrointestinal system or have been radically altered by digestive and absorption processes before the onset of illness; nevertheless, these animals displayed a reduced preference for sucrose in later tests. A number of studies with delayed illness have produced similar results.

The foregoing effects presumably have their anatomical basis in the nucleus of the fasciculus solitarius of the brain stem, where gustatory and visceral afferent fibers both terminate.

Rats are not the only animals that show genetically predisposed learning. Wilcoxon has reported (Garcia [1971]) that ". . . the bobwhite quail, which uses visual cues to identify food and to guide its pecking, is able to relate visual cues to an illness delayed for 30 minutes." Also, Grzimek [1968] of Germany has

> "... investigated the memory horses have for certain processes--in particular for the disappearance of food--by pouring oats into one of four covered boxes in front of a horse's eyes. The horse was then allowed to walk from its stand to the boxes and to eat the oats from the box which had just been filled. Even to an experienced horseman it will be incomprehensible that a horse cannot immediately grasp what seems an obvious connection. My horses went as often to the other boxes as to the one into which the food had just been put. It took very lengthy and tedious

> training to induce the animals to open first the box which had just been filled in front of their eyes. When that lesson had sunk in, the horses were no longer allowed to approach and feed immediately from the boxes which had just been filled, but had to wait for varying periods of time. One horse could keep the newly filled box in mind for only six seconds; the second horse could remember it for sixty. After a longer interval, it would again try all the other boxes.
>
> By contrast, experiments with dogs and ravens have shown that they remember for hours food which has been buried in their sight, and my wolves retained such a memory for days. It should not be concluded from such experiments that a horse has a much poorer memory in general; *only that it has a very short memory for these particular processes*. The hiding of food plays no part in the life of grazing animals, while prey often hides from the wolf, which is a hunter. It is probable that in respect of territorial recognition and dominance fighting, for example, the memory of grazers is incomparably better."

In a different vein, noting how hard it is for rats to learn to press bars to avoid shock, and for pigeons to peck keys to avoid shock (but not to get grain), we conclude that to train an animal to avoid a painful stimulus, we must choose from among its species-specific repertoire of <u>defensive</u> actions, not from its appetitive repertoire. The extensions of this principle to other realms is apparent.

Humans are not exempt from the foregoing sorts of effects. As evidence, we quote from an article by Seligman and Hager [1972].

> "Some years ago, one of us (Seligman) went out to dinner. He had an excellent *filet mignon* with *sauce Bearnaise*, his favorite, and then he went off with his wife Kerry to see the opera *Tristan und Isolde*. Some hours later he became violently ill with stomach flu and spent most of the night in utter misery. Later, when he attempted to eat *Sauce Bearnaise* again, he couldn't bear the taste of it. Just thinking about it nauseated him.
>
> At first glance, his reaction seemed to be a simple case of Pavlovian conditioning; a conditioned stimulus (the sauce) had been paired with an unconditioned stimulus (the illness), which elicited an unconditioned response (throwing up). So future encounters with the sauce caused a conditioned response (nausea). At second glance, however, he realized that the *Sauce-Bearnaise* phenomenon had violated all sorts of well-established laws:

1 The interval between tasting the sauce and throwing up was about six hours. The longest interval between two events that produce learning in the laboratory is about 30 seconds.

2 It took only one such experience for him to associate the sauce with sickness; learning rarely occurs in only one trial in the laboratory. [One shot learning].

3 Neither the *filet mignon* nor the white plate on which it was served, nor his wife, became distasteful to him; he associated none of them with the illness, only the *sauce Bearnaise*. But, according to laws of Pavlovian conditioning, all events or objects that occur along with the illness (the unconditioned stimulus) should have become unpleasant. [Specificity.]

4 His reaction had no cognitive or 'expectational' components, unlike most conditioning phenomena. When he found out that another close colleague got sick the same night, *he knew* that the sauce hadn't caused the malaise at all--stomach flu had caused it. But knowing that the sauce was not the culprit did not inhibit his aversion to it one bit.

5 Finally, his loathing of *sauce Bearnaise* stayed with him about five years, whereas associations formed by Pavlovian conditioning generally die out in about a dozen trials."

From the same article:

"... Phobias too are selective: we have phobias about heights, the dark, crowds, animals and insects, but we do not have phobias about pajamas, electric outlets, or trees, even though the latter may accompany trauma as often as the former. Isaacs Marks of London's Maudsley Hospital related a typical case. A seven-year-old girl, playing one day in the park, saw a snake, but she was not particularly alarmed. Several hours later she accidentally smashed her hand in a car door, and soon thereafter developed a fear of snakes that lasted into adulthood. Notice that she did <u>not</u> develop a car-door phobia, which would have been more logical. Moreover, considerable time elapsed between her seeing the snake and having the accident."

Turning now to mice for some other results on the nature of learning, we read from Quadagno and Banks [1973]

> "... that the procedure of cross-fostering [different species (wild and inbred) of mice pups to their respective mothers] at birth [does] not affect the ability of the cross-fostered male and female *Mus* to mate with a conspecific. [But such] cross-fostering [does affect] many other social behaviours such as allogrooming, aggressive behaviour and approach-avoidance behaviour [in the species in question]. This implies that, in [some] mice, sexual behaviour may be firmly fixed by the genotype and that it is released by the presence of a conspecific. Other social behaviours are more labile and easily changed by manipulating the early experience of these rodents."

Another early developmental result reported in the same book is that:

> "Chicks of either sex reared in partial isolation (non-contactual) establish a dominance order in a matter of hours when assembled at the age at which group-reared controls form a peck-order. These results suggest that the age at which peck-rights form is determined essentially by processes of maturation rather than of learning only..."

Another study (J. Hogan[1971]) has shown that pecking in newly hatched chicks begins as a simple response to a stimulus. Within a few weeks though, pecking is differentiated into pecking for food and simple pecking (ground scratching) in response to non-food stimuli. Presumably, later on the gizzard state and so on would also influence the latter.

On naturally predisposed habituation, Fox [1973a] reports that among

> "...certain birds, habituation seems to be involved in the differentiation of predators from friendly birds. Laboratory investigations indicate that ducklings give a crouching response to any object overhead. After a period of time, the response to object seen frequently, e.g., leaves or friendly birds, habituates. On the other hand, hawks are seen infrequently and the response never habituates. Thus as the result of selective experience, the crouching response occurs in the presence of hawks but not in response to familiar, friendly birds.
> ... Habituation could prove nonadaptive if it occurred readily to all stimuli, including stimuli that occasionally signalled danger.

> In at least some cases, animals exhibit a resistance to habituation to such stimuli. For example, sparrows fail to habituate to an owl or a model of an owl; this failure occurs even in hand-reared varieties in which other experiential variables have apparently been controlled. ..."

For another perspective, we turn next to wolves, a particularly well-studied type of social carnivore whose behavior enables us to see with special clarity the extent to which complex learning often involves comparatively simple elaborations of essentially "innate" response patterns. In the following discussion, we should note from ethology that every phenotypic feature is influenced by instructions contained in the genes, and that a behavior is "genetically determined" if it is different between individuals of different genotypes but of the same experiences (assuming the latter to be possible). A "learned" behavior is one which is different among individuals of the same genotype who have had the same experience except as regards the learned behavior. The modern ethological conception is that some behaviors are more environment-resistant, that is, less susceptible to environmental influence, than others. These others are more predisposed to be learned.

Now for some examples from Fox [1971] of highly predisposed learning in wolves. Wolves are weaned at 10 to 12 weeks, and between 6 to 12 weeks each learns on its own to catch, kill, carry, defend and eat small prey such as mice. Species typical motor patterns such as leaping, forepaw stabbing, digging, overturning feces for bugs, and shaking prey held in the mouth, are built in, and only await the appropriate stimuli for their release. What is learned are various sophistications and elaborations of these patterns. By 3 to 6 months, a wolf cub has also learned to respond properly

to its parents' moods, has learned which animals are familiar and which should be shyed from as strangers, has "imprinted" the features and terrain of its home range, and has begun to properly socialize in accordance with the dominance relations of its pack. Between one and two years, cubs learn techniques of cooperative hunting during group play such as: cutting off; ambushing; prey testing; group stalking, herding, cornering, harassing, fatiguing, mass attacking, and decoying. This learning never occurs in individuals raised in isolation, only in groups of 3 or more free to play with each other and free to trail along on adult hunts. By 3 years, wolves are fully integrated into their complex pack structures, having successfully reckoned with alpha males and females, siblings, parents, seasons, new births and deaths, and complicated individual relations.

All this reminds one of the stages of human childhood, the proverbial seven ages of man, "teen age gangs" of wild horse colts, and many other developmental parallels in the mammalian kingdom. One wonders, for example, how closely Piaget's "sets-of-sets" thinking (Piaget [1967]) that emerges at about 12 in humans is linked to the socialization stage of adolescence in all social mammals. At this stage, sets of social relations must be recognized and observed. Are there regular sequences of types of learning that are always followed in mammalian brains?

At least among humans, there is evidence that this is so in natural language learning, since children seem to learn language the same way the world over. So far, among 18 languages studied (Slobin [1972]), children progress through the same 2-, 3-, many-word grammatical stages, talking about the same subjects and rectifying the same category confusions at each stage until mature adulthood. Recent studies also suggest that chimps undergo a

similar series of cognitive developments, but have so far been taught little of man's language capability (Slobin [1972]).

Note, too, the close parallel between man and chimpanzees in spatial thinking amd memory ability. Cognitive psychologists have studied how a human can memorize a long sequence of items by imagining himself walking around a familiar building, and creating in the successive places that he passes, visual images of the things he wants to mention. This is based on the fact that human beings seem to possess specialized and powerful mechanisms for reasoning about spatial relationships. They form the basis of visual perception; Helmholtz referred to them as processes of "unconscious inference". These ideas relate to an experiment of Mentzel [1973]. He reports that

> "Juvenile chimpanzees, carried around an outdoor field and shown up to 18 randomly placed hidden foods, remembered most of the hiding places and the type of food that was in each. Their [food collection paths, taken after they were released (to retrieve the food)], approximated optimum [routes] and they rarely rechecked a place they had already emptied of food."

Obviously their spatial memory was excellent.

To add another perspective, we note that selective breeding and training in domesticated canids (expecially dogs), brings out (or by affecting thresholds, makes more accessible) certain behavioral traits which can be modified, shaped, or redirected through training. Guarding (with or without barking), herding, leading or guiding, and retrieving are examples of traits that have been bred to very low response thresholds in certain breeds. Animal trainers, of course, must exploit these predispositions.

For yet another perspective, Schneider [Anon, 1972] has found that in newborn hamsters, subcortical pathways between the tectum and the thalamus are implicated in pattern discrimination. But in adult hamsters, pattern

discrimination seems to take place in the cortex. This suggests many possible migrations of functions during ontogeny in mammalian brains. On a similar tack, Section 4 notes McLardy's position that mammalian hippocampus may be more important for survival during behavior formative period.

Still further afield, child-beating mothers are remarkably likely to have been beaten by their own mothers as children (Stoll, 1970). Here a model of behavior is "learned" in early life but is not expressed until adulthood. This reminds us of the chaffinches who were kept in auditory isolation after being caught their first summer, and who then sang the song of their species for the first time the following spring (Nottebohm [1972]). This returns us full circle to the opening topic of this Subsection; and with this we turn to the beginnings of a mathematical theory of network complexity and predisposition.

3.2. Pretheory of Network Predisposition

In Section 2.2, in our study of training a Perceptron to make binary classification, we saw that if a linear separation exists, the training algorithm will reach it. We now turn to Minsky and Papert's study of when it is indeed possible to combine the information in a given preprocessing layer to perform a given pattern recognition task. We suggest that their work be considered as one phase of a pretheory for the questions of 'predisposition' of neural networks raised in the ethological sampler of our previous subsection.

We are going to be interested in pattern predicates ψ (i.e., $\psi(X)$ is true for some patterns X, and false for others) -e.g., X is connected, or X is of odd parity, etc. - and ask such questions as, "How many inputs are required for the preprocessing modules of a one-layer Perceptron?" Of course, we can always get away with using a single element computing an arbitrary Boolean function and connect it to all the squares. So the question that really interests us is, "Can we get away with a small number of squares connected to each of the input neurons?"

Minsky and Papert show that if a predicate is unchanged by various permutations, then we may use this fact to simplify its coefficients with respect to the set of masks - and that this simplified form will often enable us to place a lower bound on the order of the predicate. Rather than giving the theory, we make this approach clear by a simple example:

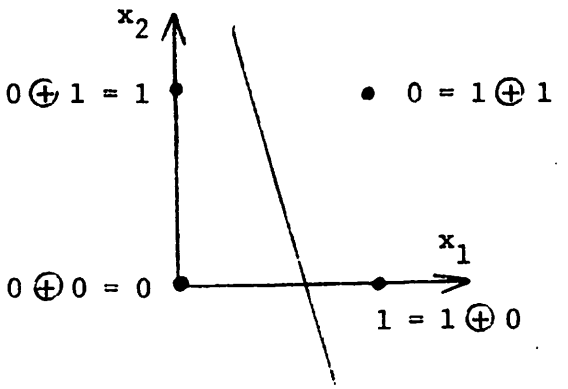

Figure 3

Consider the simple Boolean operation of addition mod 2. If we imagine the square with vertices (0,0), (0,1), (1,1), (1,0) in the Cartesian plane, with (x_1,x_2) being labelled by $x_1 \oplus x_2$, we have 0's at one diagonally opposite pair of vertices and 1's at the other diagonally opposite pair of vertices. It is clear that there is no way of interposing a straight line such that the 1's lie on one side and the 0's lie on the other side. In other words, it is clear in this case - from visual introspection - that no threshold element exists which can do the job of addition mod 2. However, let's prove it mathmematically, because in doing so, we get insight into a general technique which Minsky and Papert use over and over again.

Consider the claim that we wish to prove wrong - that there actually exists a threshold element with weights α and β such that $x_1 \oplus x_2 = 1$ if and only if $\alpha x_1 + \beta x_2$ exceeds θ. The crucial point is to notice that the function of mod 2 addition is symmetric, so that we must also have $x_1 \oplus x_2 = 1$ iff $\beta x_1 + \alpha x_2$ exceeds θ and so, adding together the two terms we have written down, we see that $x_1 \oplus x_2 = 1$ if and only if $\frac{\alpha+\beta}{2}x_1 + \frac{\alpha+\beta}{2}x_2$ exceeds θ.

Writing $\frac{\alpha+\beta}{2}$ as γ we see that, by using the symmetries of mod 2

addition, we reduce three putative parameters α, β and θ to a pair γ and θ of parameters such that $x_1 \oplus x_2 = 1$ if and only if $\gamma(x_1 + x_2)$ exceeds θ. So let's set $t = x_1 + x_2$ and look at the polynomial $\gamma t - \theta$. It is a degree 1 polynomial. Let's evaluate it at 0 where we see that we must get $\gamma t - \theta$ less than 0, evaluate it at 1 where we see that $\gamma t - \theta$ is greater than 0, and evaluate it at 2 where we must get a value less than 0. This, we see, is a contradiction. A polynomial of degree 1 cannot change sign from positive to negative more than once. We thus conclude that, in fact, there is no such polynomial, and thus we must conclude that there is no threshold element which will add modulo 2.

We now understand a general method used again and again by Minsky and Papert: Start with a pattern classification problem. Observe that certain symmetries leave it invariant. For instance, if it were the parity problem or the simple case of addition mod 2, any permutation of the points of the retina would leave the classification unchanged. We use this to cut down the number of parameters which describe the circuit. We then lump items together to get a polynomial and examine actual patterns to put a lower bound on the degree of the polynomial - and we fix things so that this degree bounds the number of inputs to the first layer of the one-layer Perceptron. For a proof of this group invariance theorem the reader may refer to Minsky and Papert's book, or to the exposition in Section 5.5 of Bobrow and Arbib [1974].

One application of the group invariance theorem shows that to tell whether or not the pattern of activated squares is connected or not requires a number that increases at least as fast as the square root of the number

of cells in the retina. Their results are most interesting, and point the way towards further insight into the functioning of the nervous system, but are restricted to highly mathematical functions, rather than the complex perceptual problems involved in the everyday life of an organism. We might note, too, that any full model of perception must not have the purely passive character of the Perceptron model, but must involve an active component in which hypothesis formation is shaped by the inner activity of the organism, and related to past and present behavior (cf. our discussion of internal models in Section 1).

We have already stressed the interest, in the study of complexity of computation, in tradeoffs between time and space. Minsky and Papert asked, "If we fix the number of layers in the network, how complicated must the elements become in order to get a successful computation?" We close this section by noting that Winograd and Spira have tackled the complementary problem of how, if we bound the number of inputs per component, we can proceed to discover how many layers of components we require. More specifically, they study algebraic functions (e.g., group multiplication), rather than look to the problem of classifying patterns. Winograd [1967] and later Spira and Arbib [1967] and Spira [1969] studied networks whose components were limited in that there was a fixed bound on the number of input lines to any component. In what follows, each module is limited to have at most r input lines. We are once again assuming a unit delay in the operation of all our modules.

The Winograd-Spira theory is based on the simple observation exemplified by Figure 4:

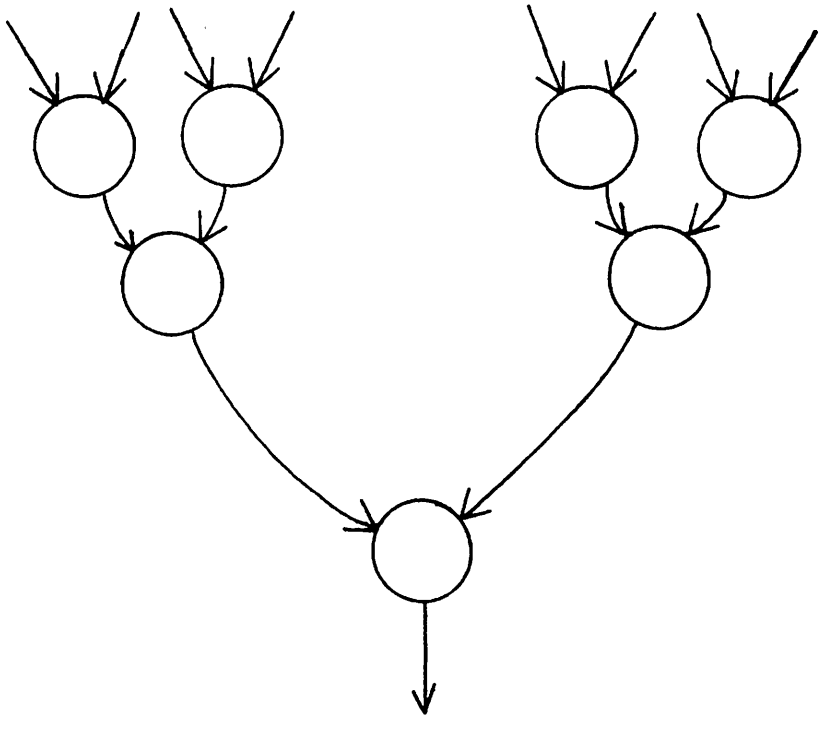

Figure 4.

Here we see that if we consider two inputs per module, then if an output line of a circuit depends on 2^3 input lines, then it takes at least three time units for an input configuration to yield its corresponding output. A lemma below formalises this observation, and is the basis for the lower bounds we obtain on computation time for various functions.

Such an apparently trivial observation has surprisingly powerful consequences for Winograd and Spira proved, for certain mathematical functions that no possible scheme of wiring neurons could compute the function in less than a certain time

delay intimately related to the structure of the function. In particular, it was shown how to go from the structure of a finite group to a minimal time for a network which computed the group multiplication. Further, Spira [1969] was able to provide for any group a network which was essentially time-optimal, in that it produced its output within one time unit of the time specified by the previously mentioned theorem.

We thus have here an extremely important result for any theory of neural networks--for a certain type of restricted component we can show how to build a network which is actually optimal with respect to the time required for computing. However, to appreciate the full complexity that lies ahead of the automata theorist who would contribute to the study of information processing in the nervous system, we must make several observations. Firstly, to achieve time optimality in his network, Spira had to use an extremely redundant encoding for the input and output to ensure that "the right information would be in the right place in the right time." The flavor of this can be given by the observation that we can multiply numbers far more quickly if they are given in prime decomposition.

$$(2^2 \cdot 3^4 \cdot 5^2 \cdot 7^1) \times (2^1 \cdot 3^0 \cdot 5^1 \cdot 7^2) = (2^3 \cdot 3^4 \cdot 5^3 \cdot 7^3)$$

than if they are given in decimal form.

This observation reminds us of the organization of the cat visual system, where a million optic fibers feed 540 million cells in the visual cortex. Thus as we move up into the visual cortex, what we reduce is not the number of channels but rather the activity of the channels, as each will respond only to more and more specific stimuli. The result is a network with many, many neurons in parallel, even though the network is rather shallow in terms of computation time. It might well be that we could save many neurons at the price of increased computation time, both by narrowing the net but increasing its depth, and also

by using feedback to allow recirculation of information for quite a long time before the correct result emerges. We see here the need for a critical investigation of the interplay between space and time in the design of networks.

The ganglion cells in the retina of a frog seem fairly well suited for an animal which lives in ponds and feeds on flies, since the brain of the animal receives specific information about the presence of food and enemies within the visual field (Lettvin, Maturana, McCulloch and Pitts [1959]) - but the price the animal pays is that it is limited in flexibility of response because its information is so directly coded. A cat (Hubel and Wiesel [1962]), on the other hand, has to process a greater amount of information to be able to find its prey - but can eat mice instead of flies. A cat cannot compute its appropriate action as quickly, perhaps, as the frog can - but makes up for that in that it has extra computational machinery which enables it to predict, and to make use of previous experience, in developing a strategy in governing its action.

We see that to completely model the behavior of the animal we must make an adequate model of its environment, and take into account structural features of the animal. It is not enough to work out an optimum network whereby a frog can locate a fly, but we must also compute whether it is optimal to couple that network to the frog's tongue, or have the frog bat the fly out of the air with its forelimb, or to have the frog jump up to catch the fly in its mouth. Clearly, the evolution of receptors, effectors and central computing machinery was completely interwoven - and it is only for simplicity of analysis that we concentrate here on the computational aspects, holding much of the environmental and effector parameters fixed. Again, we shall ignore the interesting pattern recognition problem of determining the most

effective features to be used in characterising a certain object in a given environment. For instance, to characterise a mouse one could go into many details including the placement of hairs upon its back, but for the cat it is perhaps enough to recognise a grey or brown mobile object with pointed ears and within a certain size range. It should be clear that the choice of features must depend upon the environment - if there exists a creature which meets the above prescription for a mouse but happens to be poisonous, then it will clearly be necessary for a successful species of cat to have a perceptual system which can detect features which will enable the cat to discriminate the poisonous creatures from the genuine edible mice.

We should further note that Winograd's and Spira's best results were for groups, where we can make use of the mathematical theory elaborated over the past hundred years. It will be much harder to prove equally valuable theorems about functions which are not related to classical mathematical structures. We have also made a very simple limitation on number of inputs assumption by an assumption limiting the actual types of functions that the neurons can compute, then we shall have to expect a great increase in the complexity of the theory. Some would conclude from this analysis that automata theory is irrelevant to our study of the nervous system, but we would argue that it shows how determined our study of automata theory must be before we can hope to really understand the function of the nervous system.

Another kind of learning limitation that can best be discussed in the terminology of Section 1 on interacting sets of active subsystems has been discussed by Geschwind, Luria, and others with respect to language functions in the brain. We quote a case description by Geschwind (1970) given in support of his disconnection syndrome hypothesis:

> "... this syndrome was described in 1892 by Dejerine. His patient suddenly developed a right visual field defect and lost the ability to read. He could, however, copy the words that he could not understand. He was able, moreover, to write spontaneously, although he could not read later the sentences he had written. All other aspects of his use and comprehension of language were normal. At postmortem Dejerine found that the left visual cortex had been destroyed. In addition, the posterior portion of the corpus callosum was destroyed, the part of this structure which connects the visual regions of the two hemispheres. Dejerine advanced a simple explanation. Because of the destruction of the left visual cortex, written language could reach only the right hemisphere. In order to be dealt with as language it had to be transmitted to the speech regions in the left hemisphere, but the portion of the corpus callosum necessary for this was destroyed. Thus, written language, although seen clearly, was without meaning."

Apparently, if interacting subsystems of the brain are not richly enough connected, information bottlenecks arise sufficient to make some affective knowledge ineffable and some visual sensations useless, as well as to preclude the learning of varying associations (as we documented in Section 3.1).

4. A Case Study: The Hippocampus

In Section 2, we examined studies of adaptation schemes for individual neurons; and in Section 3 we saw that different networks can learn different things, both because of the basic pattern of their internal connectivity and because of their place in the overall flux of information.

In this Section we shall review a few attempts to model some memory mechanisms in the mammalian hippocampal formation. One reason for choosing hippocampus is that more neurophysiological research on memory has been associated with it than with any other mammalian structure. Other reasons are that its circuitry is clearly organized and comparatively simple; rats and cats, our most studied mammals, have large and easily accessible hippocampi; hippocampal responses appear to correlate about as well with motor as with sensory events [O'Keefe and Dostrovsky (1971); Ranck (1973); Segal and Olds (1973); and Vinogradova (1970)], suggesting that the hippocampus plays a pivotal role in the vitally important sensory-motor relationship process; and finally, the hippocampus seems closely related to several common human diseases [McLardy (1973a, b, c, d)].

Our task in the remainder of the section will be to find memory models consistent with the neural connectivity of the hippocampus. However, we shall see that in a system as far from the periphery as the hippocampus, the significance of its input and output firing patterns is virtually unknown, which makes it hard to specify which changes in neural connectivity are "desirable"! As a result, the anatomy has proved something of a Rorschach test for neural modelers:

Kilmer and Stanley (1974) took as the point of departure for their model of dentate gyrus the general observation that successful animal behavior

must involve decisions and actions integrated in time. In order to help model such decisionary activity they considered a memory network for temporal sequences which is designed to be able to learn sequences of inputs separated by various time intervals and to repeat these sequences when cued by their initial fragments; the structure of the model being based on the anatomy of the dentate gyrus. Like the hippocampus, the model comprised a number of arrays of cells called lamellae. Each array consisted of four lines of neuromimes coupled in a regular fashion to neighbors within the array and with some randomness to neuromimes in other lamellae. The neuromimes were described by first-order differential equations. Two of the neuromime lines in each lamella were so coupled that sufficient excitation by a system input generated a wave of activity that moved slowly along the lines away from the point of excitation. Such waves effected dynamic storage of the representation via connections between lamellae. When an input was again presented to the memory, waves were excited which moved through the system as before to generate the next input's representation after the proper time interval. The system thus implemented a process of associative chaining. They plan to develop the memory circuit to allow an existing decisionary circuit to process sequences rather than single inputs, and to learn to ignore repeated nonsignificant inputs (i.e., to habituate).

Olds (1969) compared the criss-cross TA-to-GD, GD-to-CA3, and CA3-to-CA1 connections to the connection grid of a random access magnetic core computer memory. While he exploited his analogy between hippocampal circuitry and core memory arrays very well, perhaps the greatest weakness of his model was its representation of neurons by simple magnetic switches.

In Sections 4A and 4B below, we shall develop two other models, due

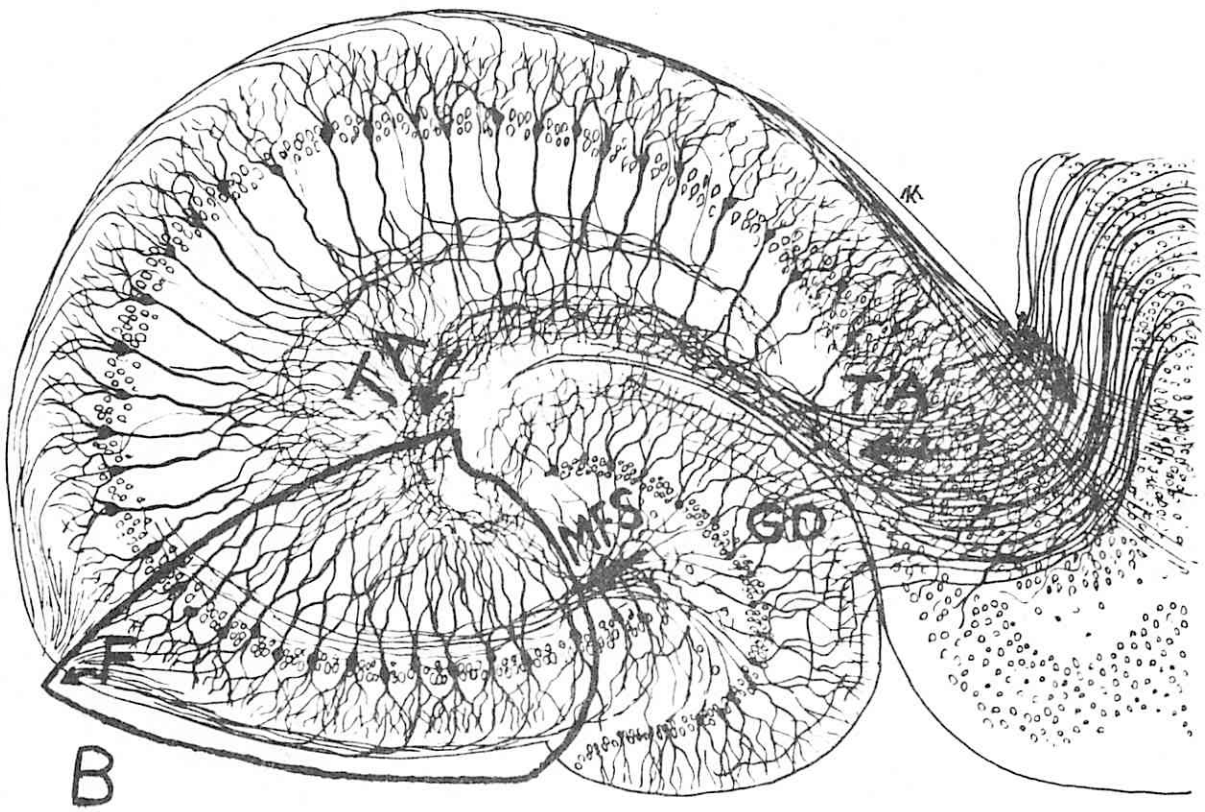

Figure 5-1

to Marr and Kilmer, respectively. It seems to us valuable to add all these schemes to the neural modeller's armamentarium, and the challenge to theorist and experimentalist alike is to design experiments which are more relevant than most to understanding the nature and mechanism of hippocampal action.

4A. The Marr Model

David Marr (1971) proposed an elegant model for the GD-CA3 subsystem of hippocampus (cf. Fig. 4-1) with the following fundamental premise: Assume that the model receives every input event E_k of a set of such events at least once and in some arbitrary order with respect to the other input events. Every E_k is presented as a binary n-tuple of 1's and 0's containing exactly M 1's. A randomly specified subset of L of these n-tuple lines feeds each codon c_i of a set of codons $c_1,\ldots,c_\ell$. The codons represent GD granule cells. Connections between input lines and codons are made through Brindley synapses, each of which contains an unmodifiable excitatory component and an excitatory component that is strengthened by each simulatenous pre- and post-synaptic activation (1 = activation, 0 = quiescence). A codon emits a 1 if the sum of its input excitations exceeds its threshold, otherwise it emits a 0. Thus an often-presented E_k will be able to excite some codons to such a degree that even a "fragment of E_k", having 0's in place of some of E_k's 1's, will fire many of them. Marr regulates his codons with supplementary inhibitory G and S cells that cause about the same number of codons to fire in response to any input, no matter what fragment of an E_k it is. The codon rank, therefore, tends to respond to a fragment of any E_k just as it first did to the most presented E_k that is sufficiently similar to it. In this sense the codon rank possesses a simple memory of all the E_k.

Randomly chosen subsets of c_i connect through Brindley synapses to each $P\Omega_j$, where $P\Omega_j$ represents a CA3 pyramidal cell. Let E_k' be any fragment of E_k. Then the function of $P\Omega_j$ is to approximately compute the conditional probability $P(\Omega_j \mid E_k')$ that the consequences of E_k', as fed into $P\Omega_j$, imply that the "completion" of the input event to E_k lies in class Ω_j. This partition is mechanized by the modifiable Brindley synapses connect-

ing the c_i to the $P\Omega_j$. The idea is that a codon rank output that fires $P\Omega_j$ hard will strengthen synapses such that similar codon rank outputs will also fire $P\Omega_j$, thereby adumbrating Ω_j. [By borrowing from Marr's neocortical model [Marr, 1970], however, a simple reinforcement signal R telling when $P\Omega_j$ is to fire can be used to alternatively define the Ω_j classes in any arbitrary way. This requires a different type of synapse between the c_i and $P\Omega_j$.]

$P\Omega_j$ approximately computes $P(\Omega_j \mid E'_k)$, which is given by a formula of the type $A[\sum_i P_i - B]/n(c_i)$, where A and B are constants, $n(c_i)$ is the number of codons firing in response to E'_k, and P_i is the probability that E'_k's completion to E_k belongs to Ω_j given that the output of c_i is 1. Marr inputs to $P\Omega_j$ an approximation to the divisive $n(c_k)$ term in $P(\Omega_j \mid E'_k)$ by adding divisively inhibitory D cells which he argues convincingly should be interpreted as basket cells. As in the codon rank, Marr also adds S and G cells to provide subtractive inhibition to regulate the level of principal output, now coming from the $P\Omega_j$. This inhibition inputs to $P\Omega_j$ an approximation to B in the $P(\Omega_j \mid E'_k)$ formula. The P_i of that formula are approximated by Brindley synaptic strengths, and A is a scale factor.

Marr's full GD-CA-3 model, if augmented with D cells in his GD portion, would contain representatives of every major cell type in GD-CA3, with nothing extra included. Note that he essentially derives the need for all but the c_i and $P\Omega_j$. His model proves that an abundance of simple, randomly connected components with plastic synapses could be put to powerful use in brains as simple classifiers of past experience.

It is rather a shame to have to omit Marr's detailed mathematics in this review, because the exact power of his results is contained in his theorems. Nevertheless, one can bypass the mathematics and still study by computer simulation all but Marr's optimality theorems and his limiting cases containing indefinitely large numbers of elements.

James Stanley did some computer simulation of Marr's model at the University of Massachusetts, and found that extremely large numbers of components would probably be needed to make the model perform at all well. While in principle this is no drawback for a brain model, in practice it makes the model difficult to refine. Some refinement seems desireable because neurons are far from codons or $P\Omega_j$ computers, and nervous connectivity is far from random.

Marr's model has enough in common with the perceptron and von der Malsburg models of Section 2 to invite comparison to both of them. Marr's codon rank corresponds to the perceptron's feature detecting layer and to von der Malsburg's E-cell layer. Note, though, that the perceptron's feature detectors do not have any inhibitory feedback for tonic regulation, but Marr's codons and von der Malsburg's E cells do. von der Malsburg's cortical cells each can be thought of as corresponding to the perceptron's output unit, and each of Marr's $P\Omega_j$ units could be converted into a perceptron-like output unit by adding a threshold and then quantizing each output signal as either 1 or 0 depending on whether $P\Omega_j$'s output exceeded this threshold or not. Marr's divisive inhibition of $P\Omega_j$ cells, which is needed to compensate for his use of Brindley synapses and large number of $P\Omega_j$ units, would still remain as an important difference; Marr's formation of Ω_j classes without the aid of external reinforcement is akin to von der Malsburg's approach, and distinct from the perceptron.

4B. THE KILMER MODEL OF CA3

Kilmer and Olinsky (1974) caricatured the known anatomy, physiology, and functional organization of CA3 hippocampus, and augmented the result with some plausible memory mechanisms. The first goal was to obtain a circuit model of CA3 that would enable us to study by computer simulation some possible CA3 memory processes involving interneurons (neurons whose axons do not leave the hippocampus proper). Because so little is known about CA3 intracellular physiology, only those processes associated with probable intercellular connectio plasticities were considered.

Fig. 4-2 outlines the connection scheme of the model, which represents only one transverse section. Extrinsic inputs to the model are interpreted as representing temporo-ammonic (TA), mossy fiber system (MFS), septo-hippocampal reinforcement (S_2), and septo-hippocampal data (S_1) inputs. The extrinsic model inputs are binary and feed pyramidal neuromimes, P_i, each of which represents a pool of perhaps a hundred pyramidal neurons of CA3. Each P_i computes, with unit latency, an output pulse repetition rate x_i which is given by

$$x_i = P_i(\mu) + P_i(E) + P_i(I) + P_i(PP) + P_i(q) + P_i(B) \qquad (1)$$

where

- $P_i(\mu)$ is a trainable nonlinear function of five non-S_2 extrinsic inputs to P_i,
- $P_i(E)$ is a sum of excitations produced by the set of excitatory interneuromimes E_j feeding P_i,
- $P_i(I)$ is a sum of inhibitions produced by the set of inhibitory interneuromimes I_k feeding P_i,
- $P_i(PP)$ is a nonlinear function of P_i-to-P_i inputs arising from functionally similar P_i units,
- $P_i(q)$ is a trainable Markovian predictive function which is positive if the probability q that x_i should be large, based on the model's response to its previous input, is high, and is negative if q is low; and

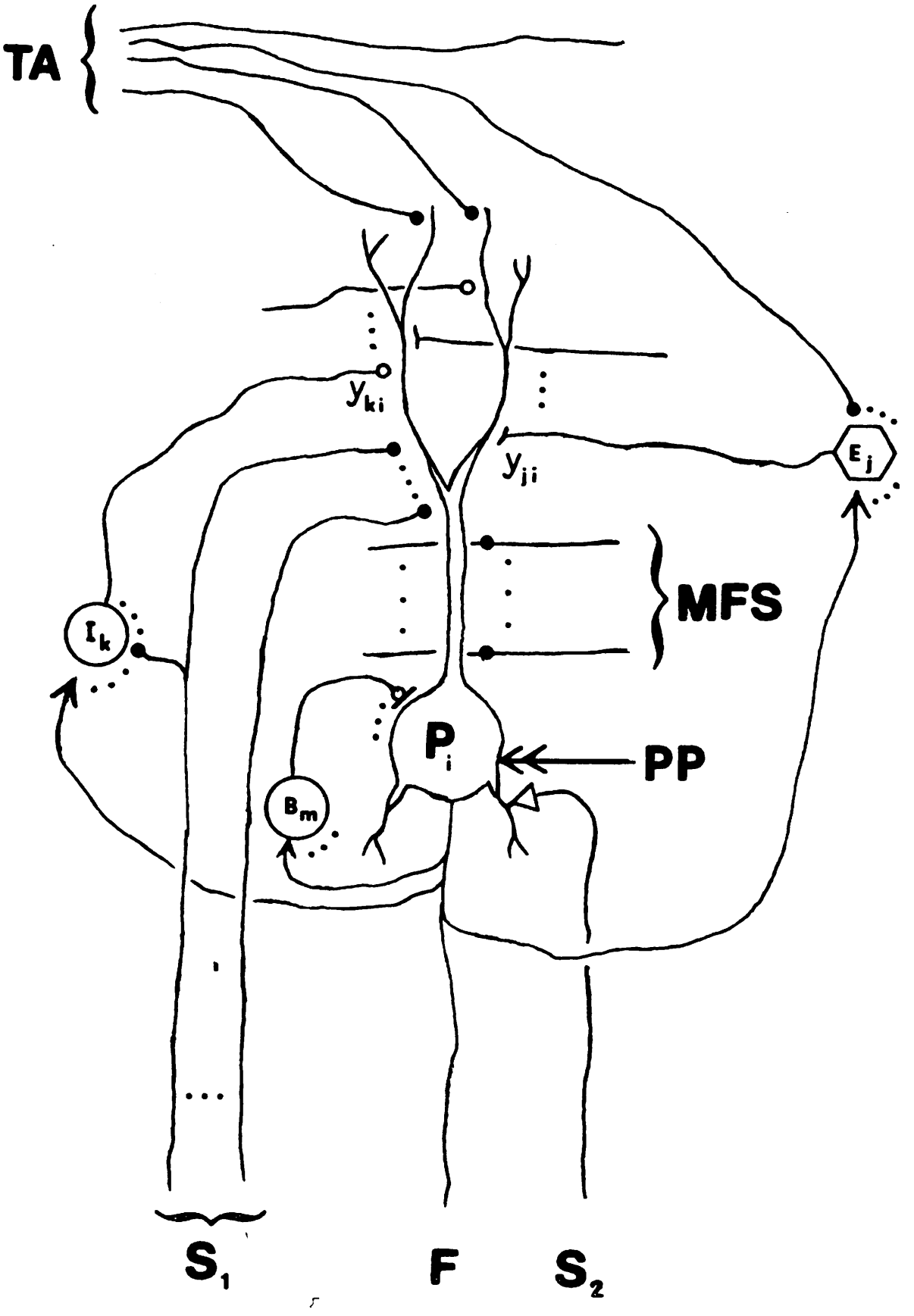

Figure 5-2

$P_i(B)$ is a divisively inhibitory basket cell interneuronal input to P_i.

These $P_i(\mu)$ functions are fully defined in (Kilmer and Olinski, 1974). Our x_i function is only meant to approximate the effects of the various kinds of influences acting on each pyramidal cell pool. The $P_i(\mu) + P_i(B)$ part of equation (1) is intended to correspond closely in effect to the $P\Omega_j$ output equation of Marr. The other terms in (1) have no counterpart in Marr's equation.

Each I_k (or E_k) in the full model (Fig. 4-3) represents a pool of perhaps 10 inhibitory (or excitatory) interneurons, whereas B_m represents a basket cell inhibitory influence on P_i which regulates by negative feedback P_i's baseline firing rate. I_k computes at each time step a nonlinear function of the inputs feeding it to produce an output burst or not, and inhibits the many P_i into which it bursts in proportion to the magnitude of the connecting synaptic strength y_{ji}. The same holds for each E_k except that it excites the P_i into which it bursts. The $y_{\ell i}$ associated with each P_i are adjusted by successive increments, on the basis of criteria signalled over S_2, in order to enable the model's output to improve with experience. The PP input to P_i represents direct pyramidal-to-pyramidal effects in CA3. F represents the fimbria-fornix output pathway of CA3. No commissural representation has been included.

All connections in Fig. 4-2 have been biologically confirmed or corroborated at least in kind, but distressingly little experimental evidence is available for magnitudes and interpretations. Single neuromimes in the model represent pools of CA3 neurons in order to better accommodate the electrophysiological data of Ranck (1973) and others. Extrinsic inputs are binary, and simply provide information on which the neuromimes of the model compute. The complete computer simulated model contains 18 P_i neuromimes; 96 each of the I_k and E_j interneuromimes; a total of 12 extrinsic input lines of the MFS, S_1, and TA types; and one S_2

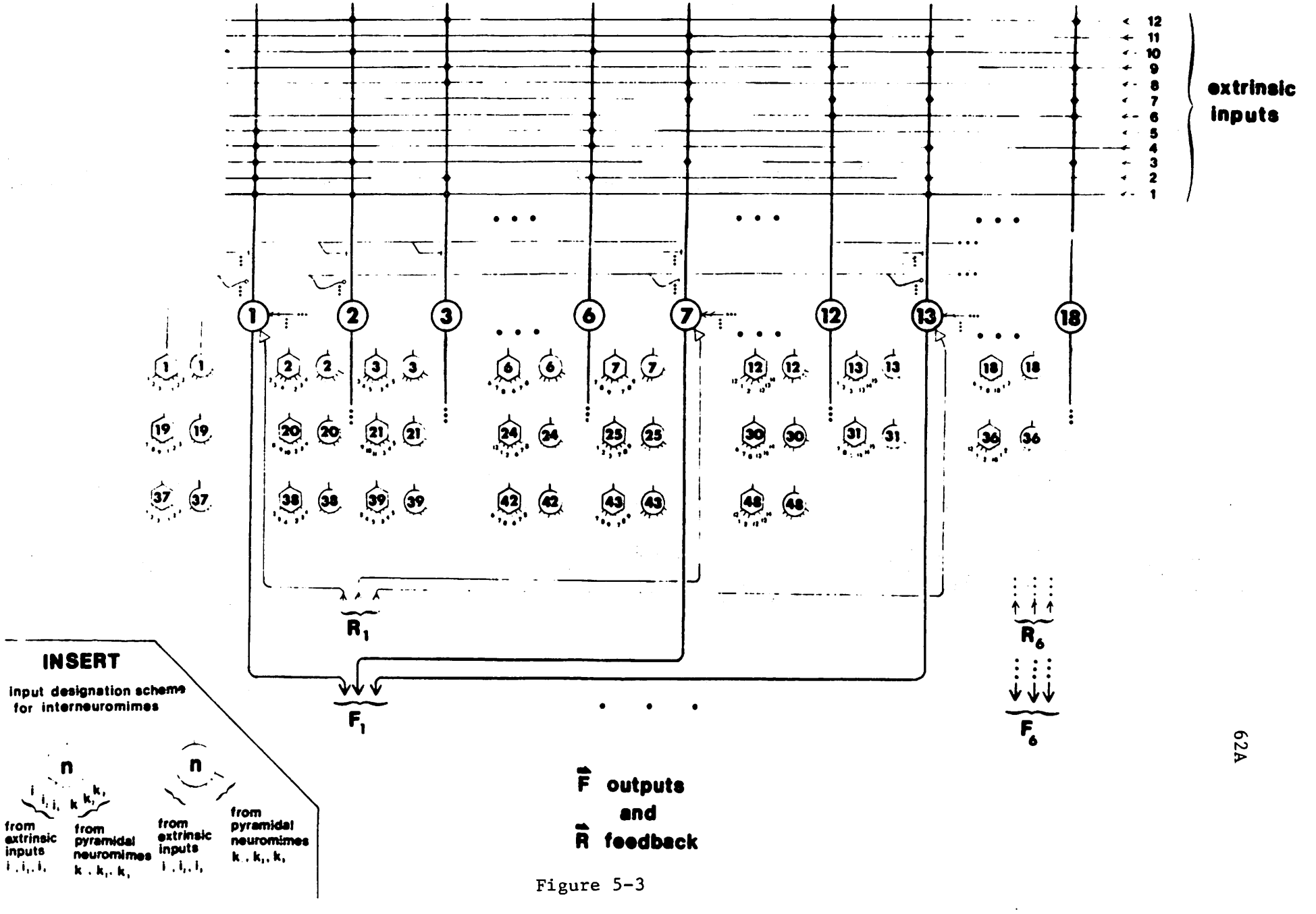
extrinsic inputs
12
11
10
9
8
7
6
5
4
3
2
1
1
2
3
6
7
12
13
18
R1
R6
F1
F6
INSERT
Input designation scheme for interneuromimes
n
i i1 i2 k k1 k2
from extrinsic inputs i . i1 . i2
from pyramidal neuromimes k . k1 . k2
n
from extrinsic inputs i . i1 . i2
from pyramidal neuromimes k . k1 . k2
F outputs
and
R feedback

Figure 5-3

line per P_i. With noted exceptions, these numbers are not intended to closely reflect corresponding proportions in the hippocampus. The resulting distortions are hopefully innocuous, and arose from our desire to keep the model as small as possible, e.g., by letting pyramidal cells "pool" their interneurons.

A fundamental idea in constructing the model was that, because interneuronal axons do not leave the hippocampus, the only direct way that interneurons can improve the quality of hippocampal output (transmitted over pyramidal axons, corresponding to F in Fig. 4-3) is to directly raise or lower pyramidal firing rates.

From a successive snapshot viewpoint of hippocampal operation, and holding all of the model's plastic mechanisms constant, the model of Fig. 4-3 operates as follows: First, an overall extrinsic input consisting of 12 binary signals in parallel is presented and fixed. Next, each P_i computes a nonlinear $P_i(\mu)$ response to the 5 extrinsic inputs it receives, where the 5 in question are randomly selected during the model's specification. This response is then modified by the addition of $P_i(PP)$, $P_i(q)$, and $P_i(B)$ quantities. Each P_i output is limited between 20 and 80 pulses per second. Next, each interneuromime computes from its extrinsic and P_i inputs a nonlinear decision to burst or not. As part of the model's specification, for each interneuron three extrinsic inputs were selected at random and three P_i inputs were counted off from a regular succession (cf. Fig. 4-3 insert). Next, the former output of each P_i is modified by the addition of $P_i(E)$, which is the sum of all those y_{ji} that are connected to I_k that are firing, and the addition of $P_i(I)$, which is minus the sum of all those y_{ki} that are connected to I_k that are firing. Next, the interneuromimes recompute their outputs in response to their new P_i inputs. This gives rise to a more or less different set of E_j and I_k firing into P_i. After several go-arounds

in the P_i-to-interneuromime-to-P_i loop, perhaps 3 or 4, an essentially stable P_i output pattern and a nearly constant interneuromime firing pattern is converged upon. These patterns are maintained until a new extrinsic input is presented to the model. A complete simulation sequences through 50 overall inputs to the model in the foregoing manner.

The goal of training the model is to ensure that for each overall extrinsic input the stable P_i output pattern is uniformly "desirable". Since input-output specifications of H have been little probed by experimentalists (Ranck's work provides an encouraging start in remedying this deficit), there is no biological evidence as to what is, indeed, "desirable" and so we simply specified ad hoc whether the "desired" P_i output for each overall input was either 20 or 80, before the model was simulated.

The full model contains two types of plasticity. The first type is associated with functions computed by the neuromimes, and the second type consists of modifiable y_{ji} excitatory and y_{ki} inhibitory synaptic strengths. We shall summarize our interpretation of these two types of plasticity in biological terms, noting that validity of the model requires only that this interpretation be approximately correct.

Kilmer and McLardy (1971) have posited, mostly on the basis of circumstantial evidence, that pyramidal cells of mammalian hippocampus are diffusely innately programmed, and that after birth, "hippocampal instincts" are quickly sharpened through experience, and then later sophisticated and elaborated in play and mimicry. The circuitry that is close to maturity at birth we have denoted "core circuitry". We envision a corresponding hippocampal process whereby at first a relatively small number of innately programmed pyramids in the core make contact with clusters of interneurons, probably under septo-hippocampal reinforcement criteria. Then other pyramids are trained--i.e., become programmed through

experience--and they in turn form cell clusters whose interneurons are partially coextensive with those in the core. Cluster after cluster is trained in this way, each to compute part-novel decisions on part-novel variations of previous hippocampal input patterns. This process diminishes in importance by post-adolescence.

Now back to the model. We assumed a prespecified core of P_i-interneuromime circuitry which had no post-developmental plasticity. The core in Fig. 4-3 consists of all the circuitry computing F_1, F_2, F_3.

After the core of the model had been determined, we wanted to train the rest so as to correspond to the developmental and learning process described above whereby cell clusters become functional one after another. We therefore grouped our 18 P_i into 6 subsets of 3 P_i each (see Fig. 4-3), and required each k^{th} subset to compute a go, no-go decision F_k as follows: if the average firing rate over the three P_i computing F_k exceeds 50, F_k = "go"; otherwise, F_k = "no-go". (We might interpret (F_1 = go) as triggering a freeze, etc.) With this F_i scheme, or course, the a priori specification of the P_j functions computing each F_i had to be identical.

We next supposed that our core circuitry computed F_1, F_2, F_3, and that a neuromime cluster-by-cluster engagement process would train F_4, F_5, and F_6 one F_i after another. A major advantage of our stage-by-stage training plan is that although each F_i is trainable using only positive and negative reinforcement, in the end the entire circuit is able to produce desired 6-dimensional output vectors $F = F_1, F_2, \ldots, F_6$.

Because animals can learn relatively arbitrary things within the predispositional constraints of Section 3, like turn left for food when a square appears and turn right when a triangle appears, criteria must be provided to any circuit that learns to select best responses from a

large repertoire to each of many different input stimuli. That CA3 can do this seems likely from the experiments of Segal and Olds (1973) and Hyden et. al. (1973), and others. That the requisite CA3 reinforcement information arises largely in the septal and medial forebrain bundle regions also seems plausible, [Stein, Wise (1971); Antelman, Lippa, and Fisher (1972); Stevens (1973); Akiskal and McKinney (1973); Ito and Olds (1971)]. We believe it likely that signals from the TA system selectively boost and depress subsets of pyramidal cells as a way of setting generalized aim, purpose, and attention information into CA3's operating program (recall that the TA system originates in entorhinal cortex [Hjorth-Simonsen (1973)], which is directly supplied by secondary association areas all over the rest of the brain [Van Hoesen, Pandya, and Butters (1972)].) Thus we interpret the R reinforcement information in Fig. 4-3 as arriving mostly over septo-hippocampal fibers in the fimbria-fornix.

We confirmed by simulation that each F_i stage could be well trained according to the above plan by simulating the following sequence of steps:

i) First, train the P_i and interneuromime functions, then train the y's

To train the y's, let each P_i's neutral output be taken as 50 pulses per second, and each desired P_i output for each overall input to the model be set a priori to either 20 or 80 pulses per second. Then, to a first approximation, an I_k can on the average improve a P_i's output by being connected to it if, over all those times that I_k fires, P_i's output should exceed 50 less often than not. In this case, the strength of the y_{ki} synapse should vary monotonically with the number of helpful minus the number of harmful such influences. Thus, if I_k fired into P_i eight times over the entire sequence of 50 overall inputs, and if for 6 of these 8 times P_i's desired output was less than 50, y_{ki} might best be made (6-2)/8 - 0.5. We have adopted a generalized version of the y_{ki} formula this example suggests. Our model's formula includes the constraint that no y_{ki} magnitude ever exceed 0.5. The main purpose of this constraint is to keep an excessive influence in those few cases where it produced an effect in the wrong

direction. Also, all y_{ki} that are minus according to a straight generalization of the formula in the example are set to 0. This is tantamount to requiring that I_k not be connected to P_i if on the average such a connection would be harmful. Every possible connection between an I_k and a P_i not prohibited by the (minus y) $\rightarrow$ 0 rule is realized with the appropriate positive y_{ki} value. Mutatis mutandis, the same conventions apply to the excitatory y_{ji} connections between the E_j and the P_i.

In the above, assume that a negative reinforcement over $\vec{R}$ implies that the F_i being trained was in error, and that a positive reinforcement implies that this F_i was correct. This assumption is not always valid, but when it is not, the effects of the error will always be corrected in step iii). In step i), interpret the training of interneuromime functions as neurological development of a statistically random set of cell functions whose only notable characteristic is that they don't fire very much.

ii) Second. To develop some terminology, suppose interneuromime I_j is firing and that it is connected to m P_i neuromimes for which k have a desired output exceeding 50 and (m-k) have a desired output less than 50. We say then that I_j "helps" in (m-k) places and "harms" in k places. Similarly, with appropriate changes, we have "helps" and "harms" for excitatory interneuromimes. In step ii) we modify interneuromime functions so that those firings in step i) that were almost as harmful as helpful (i.e., $(m-k-1) \leq k$) are deleted, and also so that all j^{th} interneuromime quiescences (non-firings) in step i) that could be transformed into firings that were helpful at every nonzero y_{ji} are so transformed. Helpful or harmful influences are recognizable at every such y_{ji} if it is always assumed that a negative reinforcement fed back over $\vec{R}$ to the currently operative part of the model (the core, plus those clusters either trained or now in training) is attributable to a mistake in the F_i currently being trained. Plausible neural mechanisms for representing the modifications in this step might easily be based on reinforcement effects at synapses.

iii) Finally, retrain the y's to correspond to the modified interneuromime functions. If an incorrect assumption about F_i's correctness were made in step i), causing F_i to be trained wrongly there, the same assumption in step iii) would cause F_i to be reversed and thereby corrected.

END OF PROCEDURE

With out three go, no-go F_i bits engaged one after another, starting with F_4, we simulated our model to determine what percent, over all 50 overall inputs to the model, of F_i outputs were correctly go or no-go as a function of the number of interneuromimes used. This was to answer the question how many interneuromimes are required in order to almost always ensure more helpful than harmful y_{ji} influences at each P_i under the best possible conditions of P_i input to interneuromimes.

In repeated simulation with varying numbers of interneurons, we found that about 200 interneurons should be connected to each P_i unit in order to reduce the model's output errors to less than 1% after training. Random interneuron functions constrained only by the requirement that they not fire very often was found to yield excellent and perhaps even best possible results. About three passes around the P_i-interneuron-P_i loop were found necessary for each model input before the model's output F_i bits stabilized.

The essential assumptions upon which our model is based are:

1) CA3 hippocampus contains numerous excitatory and inhibitory interneurons. An appreciable fraction of the inhibitory interneuronal output is provided by basket cells, and is important mainly for tonic regulation of pyramids. The remaining interneurons receive both extrinsic and pyramidal inputs in significant proportions.
2) To a reasonable first approximation, non-basket interneuronal inputs to each pyramidal pool sum nearly algebraically to influence the pool's output in direct proportion to this sum.
3) Non-basket interneurons fire infrequently over a typical day in the life of an unrestrained animal, and the relative firing patterns of these interneurons are not well correlated, at least after theta wave and other macro-rhythmic effects have been eliminated. Such random-

ness may be attributable to the way in which CA3 organizes its own impulse codes, or to factors affecting interneuronal growth and development, or to some mixture of the two.

4) The efficacy of non-basket interneuronal influences on pyramids is plastic, is equivalent to synaptic efficacy, and is established under control of an extra-hippocampal positive/negative reinforcement effect.
5) Pyramids and interneurons are drawn cluster by cluster into CA3 operation under the control of diffuse positive/negative reinforcements.

Future work will attempt to match the foregoing model's P_i output patterns to Ranck's recordings from single units in freely moving animals (Ranck, 1973) when the model is fed inputs from a habituation version of Stanley's GD model, which in turn is fed inputs from entorhinal-like units as described by Ranck. A major question will be the extent to which interneuronal feedback potentiates sequences of CA3 response instead of just momentary decisionary outputs.

KEY TO REFERENCES

Amer. Naturalist = American Naturalist

Arch. Gn. Psychiatry = Archives of General Psychiatry

Brain Res. = Brain Research

Expr. Brain Res. = Experimental Brain Research

Exp. Neurol. = Experimental Neurology

Int. J. Man-Machine Studies = International Journal of Man-Machine Studies

Int. J. Neurosci. = International Journal of Neuroscience

J. Assoc. Comp. Mach. = Journal of the Association for Computing Machinery

J. Comp. and Physiol. Psychology = Journal of Comparative and Physiological Psychology

J. Comp. Neur. = Journal of Comparative Neurology

J. Neurophysiol. = Journal of Neurophysiology

J. Physiol. = Journal of Physiology

Math. Bio. Sci. = Mathematical Bio Science

Phil. Trans. Roy. Soc. Lond. = Philosophical Transactions of the Royal Society, London.

Proc. Roy. Soc. (Lond.) = Proceedings of the Royal Society (London)

Rev. Mod. Phys. = Reviews of Modern Physics

Sci. Am. = Scientific American

REFERENCES

J.S. Akiskal and W.T. McKinney, Jr.: Depressive disorders: toward a unified hypothesis. Science, 182, (1973) 20-29.

P. Andersen and T. Lomo: Organization and Frequency Dependence of Hippocampal Inhibition. Basic Mechanisms of the Epilepsies, (1969) 604-609.

P. Andersen et al.: Lamellar organization of hippocampal excitatory pathways. Exper. Brain Res., 13, (1971) 222-238.

Anon: If you must fall on your head, do it while you're young. Monitor, New Scientist, 54, May 25, (1972) 422.

S.M. Antelman, A.S. Lippa, and A.E. Fisher: 6-hydroxydopamine, noradrenergic reward, and schizophrenia. Science, 175, (1972) 919-923.

M.A. Arbib: The Metaphorical Brain, Wiley-Interscience (1972).

M.A. Arbib: Automata Theory and Neural Models. Int. J. Man-Machine Studies, (1974) In Press.

C. Blakemore and G.F. Cooper: Development of the brain depends on the visual environment. Nature (Lond.), 228, (1970) 477-478.

C. Blakemore, J. Nachmias and P. Sutton: The Perceived Spatial Frequency Shift: Evidence for Frequency-Selective Neurons in the Human Brain. J. Physiol., 210, (1970) 727-750.

H.D. Block: The Perceptron: A Model for Brain Functioning, I. Rev. Mod. Phys., 34, (1962) 123-135.

G.S. Brindley: Nerve Net Models of Plausible Size that Perform Many Simple Learning Tasks. Proc. Roy. Soc. Lond., B. 174, (1969) 173-191.

Ramon S. y Cajal: Histologie du Systeme Nerveux de l'Homme et des Vertebres, Maloine, Paris: (1911) 2.

F.W. Cambell and J.G. Robson: Application of Fourier Analysis to the Visibility of Gratings. J. Physiol., 197, (1968) 551-566.

K.J.W. Craik: The Nature of Explanation, Cambridge University Press (1943).

R.L. Didday and M.A. Arbib: Eye Movements and Visual Perception: A 'Two Visual System' Model. COINS Technical Report 73C-9, University of Massachusetts at Amherst (1973).

M. Fox: Behavior of Wolves, Dogs and Related Canids, Harper and Row (1971).

M. Fox: Socio-infantile and Socio-sexual Signals in Canids: A Comparative and Ontogenetic Study, in M. Fox (1973) 87-115 (1973a).

M. Fox (Ed.): Readings in Ethology and Comparative Psychology (M. Fox, Ed.) Brooks Cole (1973).

D. Gabor, W.E. Kock and G.W. Stroke: Holography. Science, 173, (1971) 11-23.

J. Garcia: The Faddy Rat and Us. New Scientist, Feb. 4, 49, (1971) 254-256.

R.L. Gregory: On How So Little Information Controls So Much Behavior. Towards a Theoretical Biology 2. Sketches (C.H. Waddington, Ed) Edinburgh University Press, (1969) 236-247.

S. Grossberg: Embedding Fields: A Theory of Learning with Physiological Implications. Journal of Mathematical Psychology, 6, (1969) 209-239.

S. Grossberg: Neural Expectation: Cerebellar and Retinal Analogs of Cell Fired by Learnable or Unlearned Pattern Classes. Kybernetik, 10, (1972) 49-57.

B. Grzimek: On the Psychology of the Horse. Man and Animal: Studies in Behaviour (H. Fredrich, Ed.) St. Martin's Press (1968).

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

H.V.B. Hirsch and D.N. Spinelli: Visual Experience Modifies Distribution of Horizontally and Vertically Oriented Receptive Fields in Cats. Science, 168, (1970) 869-871.

H.V.B. Hirsch and D.N. Spinelli: Modification of the Distribution of Receptive Field Orientation During Development. Exp. Brain Res., 13, (1971) 509-527.

A. Hjorth-Simonsen: Projection of the lateral part of the entorhinal area to the hippocampus and fascia dentata. J. Comp. Neur., 146, (1973) 219-232.

J. Hogan: Development of Hunger System in Young Chicks. Behavior, 39, (1971) 127-201.

D.H. Hubel and T.N. Wiesel: Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. J. Physiol. (Lond.), 160, (1962) 106-154.

D.H. Hubel and T.N. Wiesel: The period of susceptibility to the physiological effects of unilateral eye closure in kittens. J. Physiol. (Lond.), 206, (1970), 419-436.

H. Hyden, P.W. Lange and C. Seyfried: Biochemical brain protein changes produced by selective breeding for learning in rats. Brain Res., 61, (1973) 446-451.

M. Ito and J. Olds: Unit Activity during self-stimulation behavior. J. Neurophysiol., 34, (1971) 263-273.

W.L. Kilmer and T. McLardy: A Diffusely Preprogrammed but Sharply Trainable Hippocampus Model. Int. J. Neurosci., 2, (1971) 241-248.

W.L. Kilmer and M. Olinski: Model of a Plausible Learning Scheme for CA3 Hippocampus. Kybernetik, (1974) in press.

T. Kohonen: A Class of Randomly Organized Associative Memories. Acta Polytechnica Scandinavica, E125, (1971) 19 pages.

W.J.S. Krieg: Functional neuroanatomy, Brain Books Inc., Evanston, ILL.: (1966) 3rd edition.

K.S. Lashley: Brain Mechanisms in Intelligence, University of Chicago Press (1929).

A.R. Luria: The Working Brain, Penguin Modern Psychology Texts (1973).

P. Marler: Developments in the Study of Animal Communication, in M. Fox (1973) 290-346.

D. Marr: A Theory of Cerebellar Cortex. J. Physiol., 202, (1969) 437-470; A Theory for Cerebral Cortex. Proc. Roy. Soc. Lond., B. 176, (1970) 161-234; Simple Memory: A Theory for Archicortex. Phil. Trans. Roy. Soc. Lond., 262, (1971) 23-81.

T. McLardy: Schizophrenia and temporal lobe epilepsy interrelations. IRCS International Research Comm. System, March (1973a).

T. McLardy: Habituation deficit and paucity of dentate granule-cells in some schizophrenic brains. IRCS International Research Comm. System, March (1973b).

T. McLardy: Gryus Dentatus Granule-Cell Pathology in Chronic Alcoholism. IRCS International Research Comm. System, March (1973c).

T. McLardy: Dentate Granule-Cell Sensitivity to Proximity of Blood-Vessels in Chronic Alcoholism. IRCS International Research Comm. System, September (1973d).

E.W. Mentzel: Chimpanzee Spatial Memory Organization. Science, 182, (1973) 943-945.

M.L. Minsky and S. Papert: Perceptrons: An Intorduction to Computational Geometry, Cambridge, MA: The MIT Press (1969).

N.J. Nilsson: Learning Machines, New York: McGraw-Hill (1965).

D. Noton and L. Stark: Scanpaths in Eye Movements During Pattern Perception. Science, 171, (1971a), 308-311; Eye Movements and Visual Perception. Sci. Am., 224 (2), (1971b), 34-43.

F. Nottebohm: The Origins of Vocal Learning. Amer. Naturalist, 106, (1972) 116-135.

O. O'Keefe and J. Dostrovsky: Directed Threat Response. Brain Research, 34, (1971) 171-175.

J. Olds: The Central nervous system and the reinforcement of behavior. American Psychologist, 24, (1969) 114-132.

R. Perez, L. Glass and R. Schlear: Development of Specificity in the Cat Visual Cortex. Math. Bio. Sci. (in press).

J. Piaget: Six Psychological Studies (D. Elkind, Ed.) Random House (1967).

D. Ploog: The Relevance of Natural Stimulus Patterns for Sensory Information Processes. Brain Research, 31, (1971) 353-359.

D.A. Pollen and J.H. Taylor: The Striate Cortex and the Spatial Analysis of Visual Space. The Neurosciences Third Study Program, Cambridge, Mass.: The MIT Press, (1974) 239-248.

K.H. Pribram, D.N. Spinelli and M.C. Kamback: Electrocortical Correlates of Stimulus Response and Reinforcement. Science, 157, (1967) 94-96.

K.H. Pribram: Four Rs of Remembering, in K.H. Pribram (Ed.) On the Biology of Learning. Harcourt, Brace and Jovanovich (1969) 193-225.

K.H. Pribram: How is it that Sensing so Much we can do so Little? The Neurosciences Third Study Program, Cambridge, Mass.: The MIT Press, (1974) 249-461.

M. Quadagno and E.M. Banks: The Effect of Reciprocal Cross-Fostering on the Behaviour of Two Species of Rodents, Mus Musculus and Baiomys Taylori Ater, David, in M. Fox (1973) 448-471.

J. Ranck: Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats. Exp. Neurol., 41, (1973) 461-555.

J. Ranck: Personal Communication, (1974).

F. Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms, Washington D.C.: Spartan Books, (1961).

R.C. Schank and K.M. Colby: Computer Models of Thought and Language, San Francisco: W.H. Freeman and Company, (1973).

M. Segal and J. Olds: Behavior of units in hippocampal circuit of the rat during differential classical conditioning. J. Comp. and Physiol. Psychology, 82, (1973) 195-204.

M.E.P. Seligman and J.L. Hager: The Sauce Bernaise Syndrome. Psychology Today, August (1972).

D. Slobin: Children and Language; They Learn the Same Way All Around the World. Psychology Today, May (1972) 71-82.

D.N. Spinelli: OCCAM: A Computer Model for a Content Addressable Memory in the Central Nervous System. Biology of Memory, New York: Academic Press Inc., (1970) 293-306.

D.N. Spinelli, H.V.B. Hirsch, R.W. Phelps and J. Metzler: Visual experience as a determinant of the response characteristics of cortical receptive fields in cats. Exp. Brain Res., 15, (1972) 289-304.

P.M. Spira and M.A. Arbib: Computation Times for finite groups, semigroups, and automata. Proc. IEEE 8th Ann. Symp. Switching and Automata Theory, (1967) 291-295.

P.M. Spira: The Time Required for Group Multiplication. J. Assoc. Comp. Mach., 16, (1969) 235-243.

J.M. Stanley and W.L. Kilmer: A Wave Model of Temporal Learning and Habituation in the Dentate Gyrus of the Mammalian Brain. Int. J. Man-Machine Studies, in press, (1974).

L. Stein and C.D. Wise: Possible etiology of schizophrenia: Progressive damage to the noradrenergic reward system by 6-hydroxydopamine. Science, 171, (1971) 1032-1036.

J.R. Stevens: An anatomy of schizophrenia? Arch. Gen. Psychiatry, 29, (1973) 177-189.

E. Stoll: The Abused Child. The Sciences (N.Y. Acad. Sci.), 10, (1970) 5-8 & 29-32.

N. Tinbergen: Functional Ethology and the Human Sciences. Proc. Roy. Soc., B182, (1972) 385-410.

G.W. Van Hoesen, D.N. Pandya and N. Butters: Cortical afferents to the entorhinal cortex of the rhesus monkey. Science, 175, (1972) 1471-1473.

O.S. Vinogradova: Registration of information and the limbic system. Short Term Changes in Neural Activity and Behaviour (G. Horn and R.A. Hinde, Eds.) C.U.P., (1970) 95-140.

C. von der Malsburg: Self-Organization of Orientation Sensitive Cells in the Striate Cortex. Kybernetik, 14, (1973) 85-100.

P.R. Westlake: The Possibilities of Neural Holographic Processes within the Brain. Kybernetik, 7, (1970) 129-153.

T.N. Wiesel and D.H. Hubel: Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. J. Neurophysiol., 28, (1965) 1029-1040.

D.J. Willshaw, O.P. Buneman and H.C. Longuet-Higgins: Non-holographic associative memory. Nature, 222, (1969) 960-962.

S. Winograd: On the Time Required to Perform Addition. J. Assoc. Comp. Mach., 12, (1965) 277-285.

S. Winograd: On the time required to perform multiplication. J. Assoc. Comp. Mach., 14, (1967) 793-802.

J. Zimmer: Ipsilateral Afferents to the Commissural Zone of the Fascia Dentata, Demonstrated in Decommissurated Rats by Silver Impregnation. J. Comp. Neurol., 142, (1971) 393-416.