# Two Papers on Schemas and Frames[1]

Michael A. Arbib

Computer and Information Science,
Center for Systems Neuroscience

COINS Technical Report 75C-9

(October 1975)

1. Segmentation, Schemas, and Cooperative Computation
2. Parallelism, Slides, Schemas and Frames

---

[1] The research reported in this paper was supported in part by NIH Grant No. 5 R01 NS09755-06 COM. For a development of many of the themes in this paper, see the author's forthcoming book Brain Theory and Artificial Intelligence (Academic Press, 1976).

# SEGMENTATION, SCHEMAS, AND COOPERATIVE COMPUTATION

Michael A. Arbib

Computer and Information Science
Center for Systems Neuroscience
University of Massachusetts
Amherst, MA 01002, U.S.A.

## 1. The Role of a Top-Down Approach to Brain Theory

A truly satisfying theory of any brain would place it in an evolutionary and socio-biological context. It would build upon a careful analysis of the co-evolution of the patterns of individual and social behavior which enable the animal's species to survive, and of the brain structures which enable the animal to exhibit that behavior. However, the neurophysiological and evolutionary branches of theoretical biology have been seldom conjoined. In fact, much theoretical neurophysiology can be characterized as 'bottom-up', analyzing the function of the neuron in terms of membrane properties or the behavior of small or uniformly structured neural networks in terms of simple models of neural function. I would suggest that a 'top-down' approach to brain theory can provide a bridge between 'bottom-up' neural modelling and a full-blown evolutionary and socio-biological study.

The problem is essentially this: near the periphery of the nervous system--a neuron or two in from the sensory receptors or the muscle

fibers themselves--single-cell neurophysiology allows us to make moderately useful statements about the functions of the neural networks. Thus, in these peripheral regions, the task of the neural modeller is fairly well-defined: to refine the description of the individual neurons, and to suggest missing details about their interconnections which will allow the overall network to exhibit the posited behavior. He may also, at a more abstract level, try to analyze--as, for example, Wilson and Cowan have done--the possible modes of activity of a network of a given structure. However, as we move away from the periphery, the situation becomes less clear. A given region of the brain interacts with many other regions of the brain, and it becomes increasingly hard to state unequivocally what role it plays in subserving some overall behavior of the organism. Again, the remoteness of the region from the periphery, and the multiplicity of its connections, makes it increasingly hard to determine by the methods of single-cell neurophysiology what the 'natural' patterns of afferent stimulation to that region may be. Thus, not only are we at a loss to tell what the region does on the basis of experimentation alone, but our theoretical study of modes of response of abstract networks becomes less compelling when we must expect the input to the region to be of a highly specialized kind which is unknown to us. Finally, we may expect that central regions will often be involved in 'computational bookkeeping', so that their activity will correlate poorly with stimuli or activity; and, in fact, their activity will be well nigh incomprehensible without an appropriate theory of computation.

It thus seems to me that we must complement the bottom-up approach to neural modelling of small or highly structured neural nets by what I may call <u>the 'top-down' approach to brain theory</u>: given some overall function

of the organism which is of interest to us, we must seek to analyze how that function is achieved by the cooperative computation of a number of brain regions, with the corollary specification of the natural patterns of communication between regions, and thus the specification of natural inputs for each region. It should, of course, be stressed that this theory--like any science--must succeed by successive approximations. We must start with the analysis of relatively simple functions whose operation can be approximated by the cooperative computation of relatively few regions. As simple models of this kind succeed, we may then look at more subtle descriptions of behavior, and take further account of the modulating influence of other regions. At the same time, we can expect that developments in the evolutionary and sociobiological analysis referred to above will provide us with more sophisticated descriptions of behavior with which to confront our 'top-dow-' brain theory.

While about half of this chapter will be devoted to successful brain models, the other half will be more programmatic than substantive. There are few proven methods of top-down analysis of brain function. Rather, it seems to me that further success in 'top-down' brain theory will require the injection--with substantial modification!--of many of the ideas developed in the field of artificial intelligence. This is the field in which computer scientists and cognitive psychologists have been working (sometimes together) to design computer programs which can represent knowledge, solve problems, exhibit aspects of natural language understanding, plan, etc. Attempts to truly understand the brain's higher cognitive functions may have little success without the sort of vocabulary that workers in artificial intelligence are trying to develop; but it must be stressed that workers in artificial intelligence have paid too little attention to parallelism.

In fact, of course, one of the most striking aspects of the brain is the topographical organization of its computational subsystems, and one of the major thrusts of this chapter will be to call for new concepts in a theory of cooperative computation which is adequate to handle this type of computational geography.

The aim of this chapter is not to give an exhaustive view of brain theory, but rather to exemplify the thesis of this introductory section. A more thorough review appears in Annals of Biomedical Engineering under the title "Artificial Intelligence and Brain Theory: Unities and Diversities", and, at even greater length, in my book Brain Theory and Artificial Intelligence (Academic Press, 1976).

## 2. Segmentation in Visual Perception

For many animals, a crucial part of perception is to recognize objects--in a broad sense that includes other organisms, as well as arrays of smaller objects--to the extent that the animal is able to appropriately interact with them. Thus, a full theory must model the range of activity of the organism, and analyze the perceptual clues required to extract information appropriate to this range of interaction. However, in this section I want to focus on a much simpler observation. It is clear that most environments present an animal--let us say one with a visual system--with a complex array of stimulation which is highly unlikely to come from a spatial arrangement of objects that the animal has ever encountered before. It thus becomes important to simplify the task of recognition by breaking it up into subproblems which include breaking the scene into regions--we call this task segmentation--and aggregating regions as aspects of a single object (or as jointly constituting cues for a certain course of action, etc.). Segmentation may proceed both upward from low-level cues such as depth, color, and texture; and downward from high-level cues provided by context, or overall patterns of local features.

In the remainder of this section, we present two models of segmentation: one which is neurally based and segments on depth features; and one which evolves from work on robot vision and uses segmentation on color and texture. In Section 5, we shall study a model which uses high-level semantic information to aggregate regions into collections which constitute different portions of a single object. To place these models in methodological context: the first is an example of neural modelling; the second

is in a form which suggests possible neural implementation; while the third is far indeed from a neural network implementation, and suggests, rather, the types of data structures which must be implemented in any complex perceptual system. In terms of our comments about analyzing relatively simple functions, it must be noted that each of these systems is an isolated subsystem at present. A challenge for future research is to explicate how such sybsystems can work together in a system in which depth cues, color cues, and high-level semantic cues supplement each other.

## 2A. Segmentation on Prewired Features[1]

While each retina provides only a two-dimensional map of the visual world; the two retinae between them provide information from which can be reconstructed the three-dimensional location of all unoccluded points in visual space. We indicate this in Figure 1, where the right retina can not distinguish A, B or C ($A_R = B_R = C_R$), and where the left retina can distinguish them but cannot determine where they lie along their ray. The two retinae can actually locate them on the ray: ($A_L$, $A_R$) fixes A, ($B_L$, $B_R$) fixes B, and ($C_L$, $C_R$) fixes C. There are two problems:

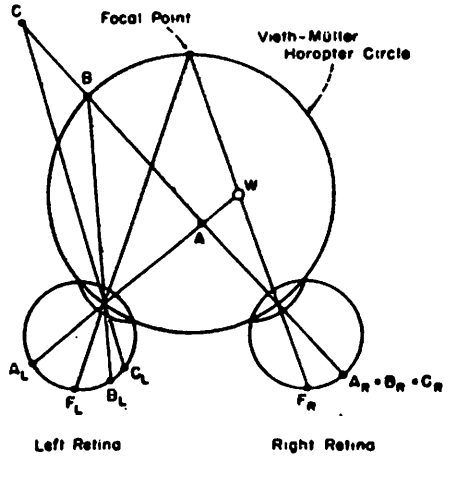



Figure 1:

The Notion of Disparity

[1] The treatment in this section follows Arbib, Boylls and Dev [1975, Sec. 2].

The first is that our observation that the two retinae contain enough information to determine the three-dimensional location of a point in no way implies that there exists a neural mechanism to use that information. However, Barlow, Blakemore and Pettigrew [1967], Pettigrew, Nikara and Pishop [1968], and others find cells in visual cortex which not only respond best to a given orientation of a line stimulus, but do so with a response which is sharply tuned to the disparity of the effect of the stimulus upon the two retinae.

The second problem is that information is given about three-dimensional location of points only when the corresponding points of activity on the retinae have been correctly paired. If the only stimuli activated in Figure 1 were at the focal point and at A, then A can only be accurately located if $A_L$ is paired with $A_R$--were $A_L$ to be paired with $F_R$, the system would 'perceive' an 'imaginary' stimulus at W.

The main thrust of the model presented below will be to suggest how disparity-detecting neurons might be connected to restrict ambiguities resulting from false correlations between pairs of retinal stimulation. But before giving the details, let us examine some psychological data which define the overall function of the model. Normal stereograms are made by photographing a scene with two cameras, with relative position roughly that of a human's two eyes. When a human views the resultant stereogram--with each eye viewing only the photograph made by the corresponding camera--he can usually fuse the two images to see the scene in depth. Julesz [1971] has invented the ingenious technique of random-dot stereograms to show,

*inter alia*, that this depth perception can arise even in the absence of the cues provided by monocular perception of familiar objects. The slide for the left eye is prepared by simply filling in, completely at random, 50% of the squares of an array. The slide for the right eye is prepared by transforming the first slide by shifting sections of the original pattern some small distance (without changing the pattern within the section) and otherwise leaving the overall pattern unchanged, save to fill in at random squares thus left blank. With Julesz's arrays, one slide presented to each eye, subjects start by perceiving visual 'noise' but eventually come to perceive the 'noise' as played out on surfaces at differing distances in space corresponding to the differing disparities of the noise patterns which constitute them.

Note well that both stimuli of the stereogram pair are random patterns. Interesting information is only contained in the correlations between the two--the fact that substantial regions of one slide are identical, save for their location, with regions of the other slide. Then the visual system is able to detect these correlations. If the correlations involve many regions of differing disparities, the subject may take seconds to perceive so complex a stereogram--during which time the subjective reports will be of periods in which no change is perceived followed by the sudden emergence of yet another surface from the undifferentiated noise.

To clarify the ambiguity of disparity in Julesz stereograms, let us caracature the rectangular arrays by the linear arrays of Figure 2. The top line shows the 21 randomly generated 0's and 1's which constitute the 'left eye input', while the second line is the 'right eye input' obtained by displacing bits 7 through 13 two places left (so that the bit at i position goes to position i - 2 for $7 \leq i \leq 13$) while the bits at position 12 and 13

thus left vacant are filled in at random (in this case, the new bits equal the old bits--and event with probability 1/4), with all other bits left unchanged. Then in the remaining 5 lines of the figure we show a disparity array, with the $i^{th}$ bit of the disparity of line D being a 1 if and only if the $i^{th}$ bit of the 'right eye input' equals the $(i+d)^{th}$ bit of the left eye input.

The disparity array of Figure 2 suggests the stripped-down caricature of visual cortex which we shall use for our model. Rather than mimic a columnar organization, we segregate our mock cortex into layers, with the initial activity of a cell in position i of layer d corresponding to the presence or absence of a match for the activity of cell i of the right 'retina' and cell i + d of the left retina. (This positioning of the elements aids our conceptualization. It is not the positioning of neurons that should be subject to experimental test, but rather the relationships that we shall posit between them.) As we see in Figure 2, the initial activity in these layers not only signals the 'true' correlations (A signals the central 'surface'; B and D signal the 'background'), but we also see 'spurious signals' (the clumps of activity at C and E in addition to the scattered 1's, resulting from the probability of 1/2 that a random pair of bits will agree) which obscure the 'true' correlations.

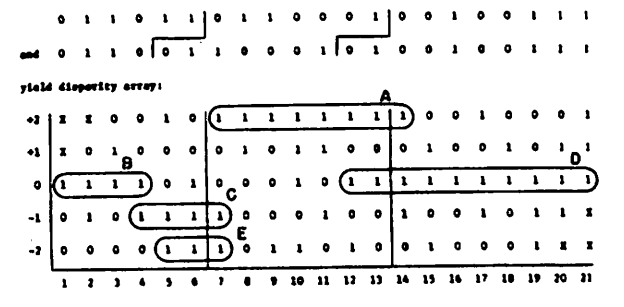



Figure 2:

Segmentation on Disparity Cues

Let us now place this in a more general context (Dev [1975]), in which we have any set of prewired features--with one spatially coded array of detectors for each feature. We then have the following situation for the problem of segmentation of prewired features:

Conceptualization: 'Layers' of cells (they are really in 'columns'), one for each prewired feature.

Principle: Minimize the number of connected regions.

Possible Solution: Moderate local cross-excitation within layers; increasing inhibition between layers as difference in feature increases.

Can we, then, interconnect the 'layers' in such a way that clearly defined segments will form? We might imagine (but only as a crude first approximation) the resultant array of activity as then providing suitable input for a higher-level pattern-recognition device which can in some sense recognize the three-dimensional object whose visible surfaces have been so clearly represented in the brain.

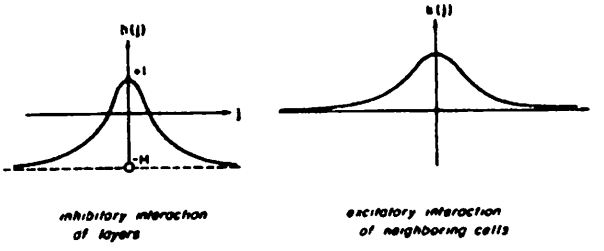
While we do not have a precise functional 'region measure' which we have minimized, we do have a plausible interconnection scheme which yields qualitatively appropriate behavior of the array: The essential idea is given by the rule that there be moderate local cross-excitation within a layer; and inhibition between layers which increases as the difference in feature increases. Let then $x_{di}(t)$ represent the activity of the cell in position i of layer d at time t (where we now let activity vary continuously between 0 and 1); and let $h(j)$ and $k(j)$ be functions of the form indicated by Figure 3. Then the change of activity of a cell is given in our model by the equation

$$(1) \qquad x_{di}(t+1) = \sum_{d'i'} \sum h(d-d')k(i-i')x_{d'i'}(t) + x_{di}(t_0)$$

where it is understood that the sum 'saturates' at 0 and at 1.

Figure 3:

Interaction Coefficients




What this scheme does is allow a clump of active cells in one layer to 'gang up' on cells with scattered activity in the same region but in other layers, while at the same time recruiting moderately active cells which are nearby in their own layer. The system then tends to a condition in which the activity is clearly separated into 'regions', with each region having its own unique feature (layer of activity). In other words, such a scheme resolves feature ambiguity through suppression of scattered activity, thus permitting activity related to only one feature in any one locale. Moreover, returning to the stereopsis example, the dynamics of the model does represent the Julesz phenomenon of a noise stereogram taking some time to be perceived, with each new surface being perceived rather abruptly. This is simulated in the model by the fact that, once a sufficient number of clumps achieve high activity, the recruiting effect fills in the gaps between the clumps to form a good approximation to its final extent.

Before closing our discussion of this model, we should note that equation (1) can be rewritten in a fashion which suggests a plausible scheme of neural interconnection.

We decree that the neurons of the above array all be excitatory. We now introduce a layer of inhibitory interneurons, one for each spatial direction, the $i^{th}$ of which has activity at time t given by the simple equation

$$y_i(t) = \sum_d x_{di}(t).$$

Let us now pick a constant H such that $\bar{h}(j) = H + h(j) \geq 0$ for all j. We may then rewrite (1) in the form

$$(2) \qquad x_{di}(t+1) = \sum_{d'} \sum_{i'} (\bar{h}(d-d') - H)k(i-i')x_{d'i'}(t) + x_{di}(t_0)$$

so that

$$(3) \qquad x_{di}(t+1) = \{ \sum_{d'} \sum_{i'} \bar{h}(d-d')k(i-i')x_{d'i'}(t) \} - \sum_{i'} l(i-i')y_{i'}(t) + x_{di}(t_0)$$

where $l(i-i') = Hk(i-i')$. Thus (3) shows that our model may be given structural expression in a form in which the $x_{di}$ are all excitatory, with excitation being appropriately counteracted by inhibition from single layer of inhibitory interneurons.

## 2B. Segmentation on Ad-Hoc Features

While anyone who has used the focus control of a camera finds it plausible that a small number of different disparities can give a tolerable set of depth cues to aid other mechanisms for locating objects in the world, credulity would be strained by the suggestion that we have a prewired set of features for every color or texture which will prove of value in setting off one region of the visual world from another. In this section, then, we present a scheme which creates ad hoc features for segmenting visual input. It is due to Hanson, Riseman and Nagin [1975], who give full references to related literature. The scheme is part of a preprocessor for the visual system of a robot which is to analyze outdoor scenes. In what follows, I describe a scheme of segmentation based on their model, rather than detailing their computer implementation. Some of the computations in the scheme have clear neural implementations; other parts challenge us to find neural implementations.

The system input consists of three spatially coded intensity arrays, one each for the red, green and blue components of the visual input. The first task of the system is to extract microtextures--features, such as hue, which describe small 'windows' in the scene. Even in the foliage of a single tree, or in a patch of clear blue sky, the hue will change from window to window, and the system must be able to recognize the commonality amongst the variations. However, in segmenting a natural scene, macrotexture will be more important than microtexture. Macrotexture is a pattern of repetition of (one or more) microtextures across many local windows, and its recognition requires the analysis of structural relationships between types of microtexture. For example, the branching of a tree would have the microtexture of leaves interspersed with that of shadows in summer; while the microtextures of branches and of sky might characterize its winter appearance.

We extract microtexture first. The aim is to do this without using predefined features. The general method is as follows: pick n feature parameters; map each image point into the feature space forming an n-dimensional histogram. Then apply a clustering algorithm to segregate the points into a small number of clusters in feature space. Each cluster then forms a candidate for a microtexture of use in segmenting the original image. The point is that, for example, while tree foliage and grass may each yield a range of greens, the feature points of the two regions should form two clusters with relatively little overlap. The result of the clustering soperation is suggested by the (hypothetical) example of Figure 4.

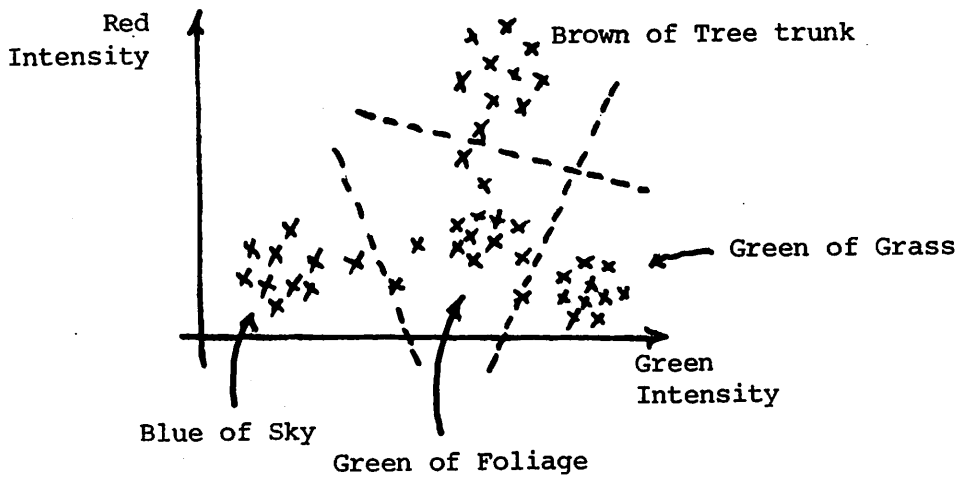



Figure 4.

An image of a tree in a field is mapped into feature space. The four microtextures of blue of sky, green of foliage, green of grass, and brown of tree trunk define four clusters in feature space. The lines drawn to separate clusters are somewhat arbitrary, and further processing is required to settle 'demarcation disputes'.

Once the clusters have been formed, each may be assigned a distinct label. Returning to the original image, a cluster label may be associated with each point, which thus has a tentative, and ad hoc, microtexture associated with it. So far, however, no spatial information has been used to bind texture elements together. If N clusters had been formed, spatial information is used to constrict an N by N adjacency matrix, in which the (i, j) element records. The number of times a point labelled i is adjacent to a point labelled j (i.e. bearing the microtexture label of the $j^{th}$ cluster). A homogeneous region of points from a single cluster j will yield large values of the (j, j) element of the matrix. A large number of adjacencies between points of two cluster types may signal a cohesive region that has a repetitive mixture of cluster types--in which case it determines a microtexture. [However, note that two microtextures may be interdistributed in different ways to determine macrotextures.]

Macrotextures suggested by large (i,j) entries can then be used as labels for the final pass of region-growing on the image.

The contribution from a boundary between two homogeneneous regions of types i and j, respectively, would distort the (i,j) entry. To avoid this, it pays to remove large connected homogeneous regions, and then form a modified adjacency matrix without these boundary contributions. Another boundary-value problem (!) is that clustering may yield a cluster due to windows overlapping 2 regions. However, if we apply curve-following, as well as region-growing, algorithms to the cluster types, we can generate boundaries directly to supplement our region-growing process.

The resultant regions, labelled by macrotexture features, can then provide the input to a semantic labelling process of the kind we shall discuss in Section 5.

## 3. Competition and Cooperation in Neural Networks

Selfridge [1959] posited a character recognition system, called Pandemonium, which would behave as if there were a number of different 'demons' sampling the input. Each demon was an expert in recognizing a particular classification and would yell out the strength of its conviction. An executive demon would then decree that the input belonged to the class of whichever demon heard yelling the loudest.

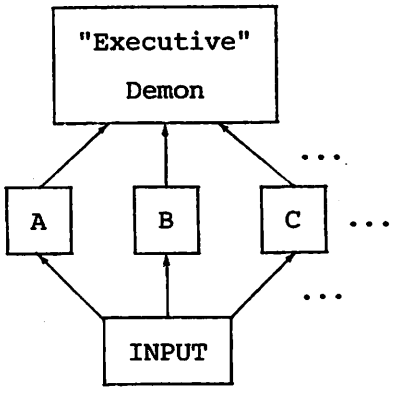



Figure 5.

Pandemonium

On the other hand, Kilmer, McCulloch and Blum [1969], in modelling the reticular formation, posited a system without executive control. Rather, each of an array of modules sampled the input and made a preliminary decision to the relative weights of different modes as being appropriate to the overall commitment of the organism. The modules were then coupled in a back-and-forth fashion so that eventually a majority of the modules would agree on the appropriate mode--at which stage the system would be committed to action. A reasonable analogy is a panel of physicians sharing symptoms and coming to a consensus about a diagnosis for a patient. (This suggests that

social analogies may once again play an important role in brain theory.)

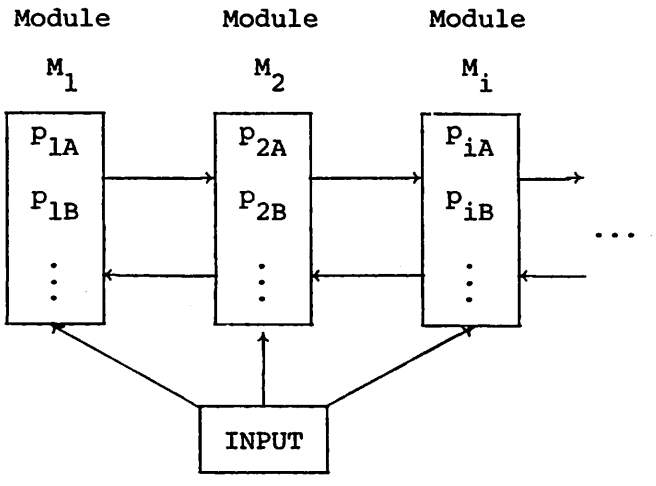



Figure 6.

S-RETIC

Here is the operation of the model, S-RETIC, in more detail. Each module $M_i$ receives a sample $\gamma_i$ of the input lines, with more nearby modules tending to receive more highly overlapping samples. At each time period, $M_i$ emits as output a probability vector $p_i = (p_{i1},\ldots,p_{im})$, where $p_{ij}$ is the weight that $M_1$ currently assigns to the hypothesis that the system input justifies the system committing itself to mode j of overall activity.

In addition to $\gamma_i$, $M_i$ receives 2 input vectors $p_i^H$ and $p_i^L$. Each $p_{ij}^H$ is $p_{kj}$ for some $k > i$ depending on i and j; similarly $p_{ij}^L$ is a $p_{kj}$ for some $k < i$. The k's are distributed randomly in such a way that small values of $|k - i|$ are favored.

A transformation $\Gamma_i$ transforms the sample $\gamma_i$ into a probability vector $p_i' = \Gamma_i(\gamma_i)$ --the best mode estimate on the base of the sample $\gamma_i$ alone. $p_i^H$ and $p_i^L$ are normalized to yield probability vectors $p_i'' = N(p_i^H)$ and $p_i''' = N(p_i^L)$.

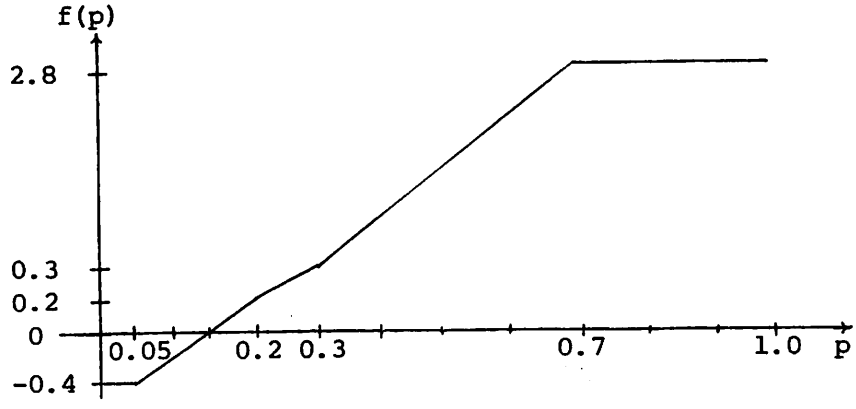



Figure 6:

This is the curve for m = 4: p = 0.25 is then the region of minimal information.

Each of the components of these 3 vectors is then operated upon by the nonlinear function f which accentuates p-values above 0.3 and diminishes those below 0.2. f implements 'redundancy of potential command'--those modules with the most information about a mode have most authority.

A secondary mode estimate is then formed by the formula

$$\frac{C_\pi f(p_y') + C_\delta f(p_y'') + C_\alpha f(p_y''')}{C_\pi + C_\delta + C_\alpha}$$

where the $C_\pi$, $C_\delta$ and $C_\alpha$ are adjustable weights, such that $C_\pi$ is far greater than $C_\delta$ and $C_\delta$ at times of S-RETIC change; but relax to roughly equal values thereafter. [Note the computation does not depend on the old output vector of $M_i$--it is as if $M_i$ "forgets" what it has learnt unless somebody later "reminds" him of it.] The vector $\bar{p}_i$ is then passed through a transform R to form the output vector $p_i$. R operates by accentuating differences between small components, and shrinking between large components; adding a scalar to all components to make them positive; reversing the first process; and then normalizing: $R = N \cdot \bar{h}^{-1} \cdot T \cdot h$.

We say S-RETIC converges to output mode j if more than 50% of the modules indicate the $j^{th}$ mode with probability > 0.5. In computer simulations,

convergence always took place in less than 25 cycles; and, once converged, stayed converged for that input. Strong $p_{ij}$ values for a given j are more likely to switch the net into mode j if the i's are close than if the i's are widely scattered.

Didday [1970], in modelling the snapping behavior of a frog confronted with two flies, posited a system of competitive interaction in the frog's tectum, which would lead in most cases to the suppression of all but one region of 'bugness' signalling, and result in the frog's snapping at one of the flies which caused the visual stimulation. In some cases, however, no region would emerge victorious from the competition. More precisely, a frog confronted with several wiggling stimuli (be they flies, or the motion of the tip of the experimenter's pencil) may exhibit one of three responses:

(a) Snap at one of the stimuli

(b) Snap at the 'average position' of two stimuli

(c) Snap at none of the stimuli.

On the basis of the Pitts-McCulloch model [1947] of superior colliculus, and the Braitenberg-Onesto model [1960] of cerebellum, Arbib [1972, Sec. 5.5] posited the existence of a distributed motor controller which responds to a pattern of high-level stimulations on its input surface by triggering a motion to the center of gravity of the spatial positions encoded by the loci of high-level inputs. [Regrettably, experimental data on type-coding vs. target-coding of actions in mammalian brain is still very indecisive.] Such a controller must be hierarchical since, for example, with increasing angle, a frog's head movement requires activity mainly in neck muscles, then trunk involvement, until with greatest angles hind-leg stepping is required. The above considerations lead to the scheme of Figure 7.

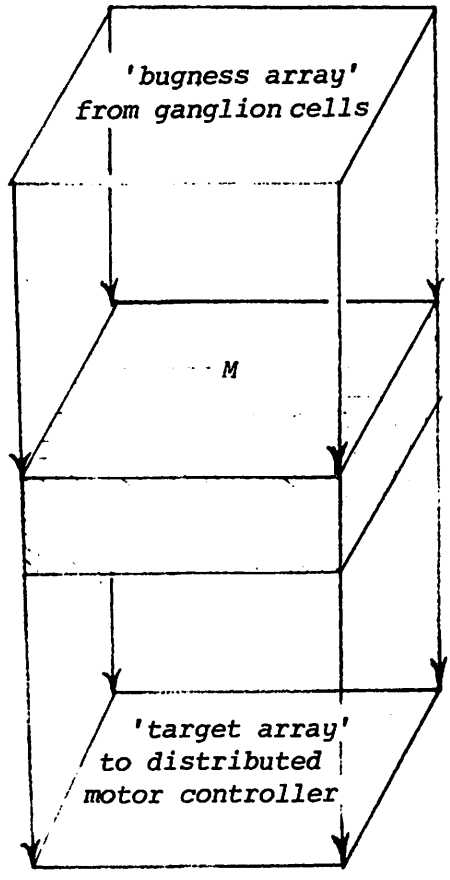



Figure 7

Our task is to give a structural description of the box M whose input-output behavior is suggested by the experiments. Namely, given a spatial input array I(x,y) with well-defined peaks at $(x_1,y_1),\ldots,(x_n,y_n)$, say, M should emit an output array such that:

(a) Normally, the output array will have a single large peak at the $(x_j,y_j)$ for which $I(x_j,y_j)$ is maximal.

(b) Occasionally, the output array will have two peaks corresponding to the two largest $I(x_j,y_j)$ --the motor controller will convert this into a snap at the average position.

(c) No part of the output array will reach the trigger-level for the motor controller.

In short, M resolves the redundancy of potential commands in $I(x_1,y_1),\ldots,I(x_n,y_n)$ to enable the frog to snap, normally, at one food-worthy location. Even case (b) may be valuable, as when a frog snaps between two wiggles which are the ends of a worm. The goal was to keep the logic distributed rather than channeled through the serial computation of a localized executive. This, in fact, Didday achieved using two layers of cells, whose names suggest their relation to cell types actually observed by Lettvin *et al.* [1960]:

(i) The sameness cells: sum the total 'bugness' activity outside their own region.

(ii) The newness cells: signal change in 'bugness' in an area.

More specifically, let

$f(x,y,t)$ = the 'bugness' evaluation made by the tectum on the basis of ganglion cell activity; at position x,y; at time t.

$m(x,y,t)$ = a 'masked' evaluation of bugness--the output of M at position x,y; at time t.

$s(x,y,t)$ = the activity of the sameness cell at position x,y; at time t.

$n(x,y,t)$ = the activity of the newness cell at position x,y; at time t.

These activities are then related as follows:

$$s(x,y,t+1) = \Big[\sum_{x',y' \notin B_{x,y}} m(x',y',t)\Big] \Big/ \Big[1 + \sum_{x',y' \in B_{x,y}} m(x',y',t)\Big]$$

where $B_{x,y}$ is a small region around (x,y)--i.e., s(x,y) is large to the extent that most m activity is outside $B_{x,y}$.

$$m(x,y,t+1) = n(x,y,t+1) + \{m(x,y,t) \Big/ [1 + h(s(x,y,t+1))]\}$$

$$\text{where } h(s) = \begin{cases} 0 & \text{if } s < .2 \\ s & \text{if } .2 < s \leq 1.6 \\ 2s-1.6 & \text{if } s \geq 1.6 \end{cases}$$

so that m remains unchanged where $B_{x,y}$ contains a large proportion of the 'masked bugness'; and is drastically reduced where $B_{x,y}$ contains very little

while

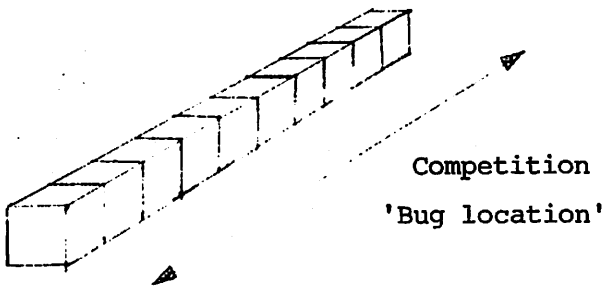
n(x,y,t+1) = increase in f(x,y,t) over f(x,y,t-1) if this is positive unless the cell is habituated--a complication we ignore here.

Thus in the nonhabituated frog a new pattern is entered into the masked cell layer; thereafter (for constant input) the distributed computation iterated through the sameness cells suppresses all but the highest peak of input activity as represented in the masked cell layer. The drawback with this model is that the initial values entered by the newness cells are larger than the final values to which the m(x,y,t) converge, so that one must assume that newness activity inhibits any motor effect of m(x,y,t) until after a convergence period. Normally the converged pattern has only one sharp peak--whether or not it is above threshold determines whether case (a) or (c) obtains. In some cases, two nearby peaks in f coalesce in m, and we have case (b).

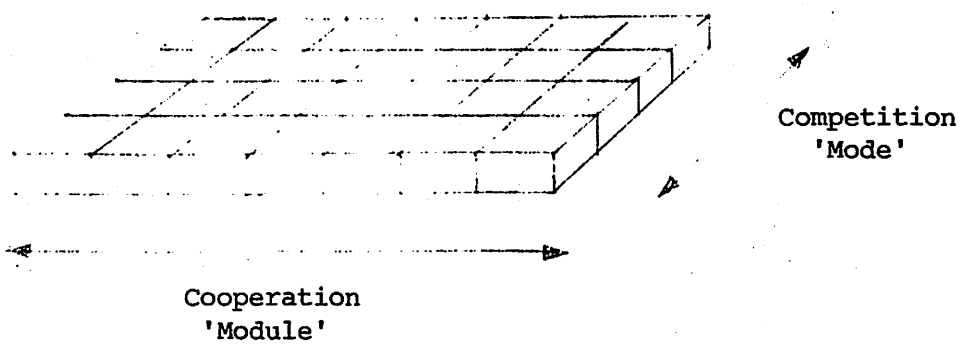
In attempting to place these studies in perspective, Montalvo [1975] observed that we could analyze the Dev, Didday and S-RETIC models within a common framework, with the computational subsystems arrayed along two dimensions, one of competition and one of cooperation, as in Figure 8.

The theme of competition and cooperation has thus emerged in three completely separate neural network models. As we shall see in the next two sections, it also plays a role when we look at the way in which an internal model of the world would operate.

(a) Didday




(b) Kilmer and McCulloch




(c) Dev

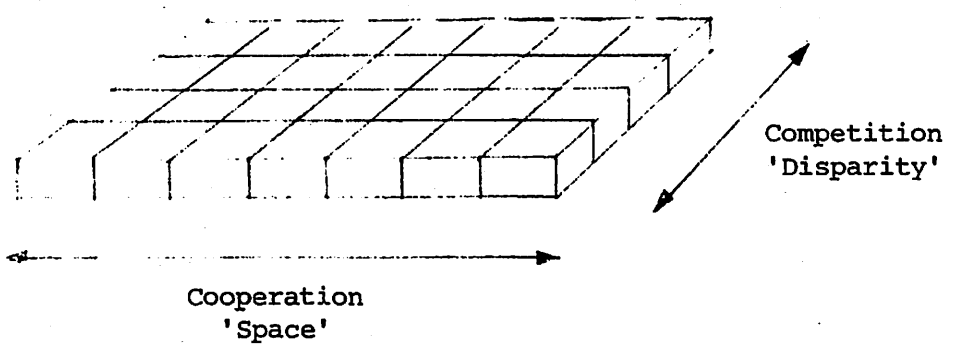



Figure 8

Competition and Cooperation in three neural nets.

## 4. Representing the World as an Array of Active, Tuneable Schemas

We have stressed that an organism's sensory stimuli should be analyzed as an array of familiar 'objects'--whether the object was a single object, such as a tree, or a composite object such as a row-of-trees. Moreover, the representation of this array should be easily updateable. This 'array representation' is reminiscent of the making of movie cartoons, where each frame is obtained by photographing the contents of a 'slide-box'--in which are appropriately positioned slides. A background slide may remain the same for many successive frames; a middle-ground slide (representing, say, a tree) may require no redrawing, though it may require repositioning relative to the background; whereas a foreground slide (representing, say, a person) may remain the same in gross features from frame to frame yet require constant redrawing of details (to represent, for example, the movements of mouth and limbs).

The basic slide-box metaphor (Arbib [1970; 1972, p. 92]), then, requires the representation of input patterns to be given as an array of slides--appropriately located and 'tuned'--chosen from a slide-file, a collection of standard slides, with the array 'covering', in some suitable sense, the input. However, rather than thinking of slides in visual terms we must regard them as cueable by various modalities--the slides are to cover the 'sensible environment'--so that a cat-slide may be activated as much to cover a meow as to cover the sight of a cat. More importantly, though, from an action-oriented point of view, the activation of a slide is now to be seen as giving access to programs for guiding possible interaction of the organism (be it human, animal or robot) with the 'object' which the slide represents.

The important point is that we wish to analyze scenes which extend over time, and in the slide-box representation a frame can often be obtained from its predecessor by simply relocating the 'active' slides (i.e., those currently in the slide-box) and updating some of their parameters--rather than resegmenting the input and assigning tags to these regions. Of course, when some region is no longer tolerably covered by the active slides, some appropriate retrieval mechanism must replace one or more slides by some new slide from the slide-file, and tune it appropriately.

But this language smacks too much of the metaphor. Let us replace slide by the notion of a schema--an array of programs to analyze a segment of the input to determine a possible course of action. As such, a schema must be relocatable, tuneable, and linkable with other schemas. Thus, we have the following components of a schema:

(1) Input-Matching Routines: A routine which succeeds to the extent that it covers a spatial region of multi-modal input (so that a cat schema can cover a furry region which emits meows, but not one that emits barks). This can be biased by non-sensory context inputs. The input-matching routines may include calls for confirming information (as in the eye-movement calls of Didday and Arbib [1975]). The level of success of these input-matching routines may be regarded as an 'activation' level of the schema. The activation level of the schema increases as the location and other parameters of the schema are adjusted to better fit the covered region. However, to complicate the story, the resolution level (i.e., the precision of this parameter match) required for activation to saturate may well depend on goal settings, or other non-sensory input to the schema--consider the different level of precision with which we 'perceive' a tree when we intend to walk around it, as against when

we intend to climb it (so that the placement and estimated strength of the branches, rather than the general area they occupy becomes important).

(ii) Action Routines: To the extent that the success of the input-matching routines raises the activation level of a schema to that extent do certain actions become appropriate for the organism. Programs for some of these actions then form part of the schema. A crucial integrative property of schemas is that increasing accuracy of parameter adjustment by the input-matching routines automatically adjusts parameters in the action routines in such a way that the action becomes more appropriate for the current environment and goal structure (if the schema has been properly 'evolved'). A classic example is that the cat turning its head to gaze at a mouse is automatically tuning its motor system for the pounce.

(iii) Competition and Cooperation Routines: To date, we have talked of a schema as acting in isolation, attempting to raise its activation level by proper matching of input. As we shall discuss in the next section, the operation of competition and cooperation routines helps determine which population of parameter-adjusted highly active schemas will constitute the current model of the environment. Of course, there still remains the problem of determining which of the action routines of these schemas are to operate--and another network of competition and cooperation routines will be involved in determining a compatible set of actions. [This expository separation--scene analysis first, planning second--is misleading. At any time, the organism is engaged in some activity, even if that be resting. Thus it is not so much a matter of choosing a course of action as it is

of determining whether the time has come to change the course of action being pursued. The completion of an action may remove it from the competition. More interestingly, the execution of an action may provide new sensory input which de-activates the slide (or drastically changes the parameter setting) which enabled the action--as when we bite into what appears to be a piece of fruit only to discover that it is made of wax.

Notice that all these routines provide the semantics of a schema--what an object means to us comprises our knowledge of what we can do with the object and what relations it has with the object, to the extent that our input-matching routines can capture the effects of these actions and relationships. In any case, we have come a long way from the original notion of a slide as being simply a coloured transparency that approximates a region of the visual field. Now that we have our new general notion of a schema we shall henceforth reserve the word for its technical sense of an (input-matching; action; cooperation and competition) set of routines, with the crucial relationship between the parameters of the input-matching and action routines.

We must now look more carefully at the structure of parameter sets. Minsky [1975] notes that a person cannot visualize a cube in any perspective with great accuracy. However, I do not accept his suggestion that we can internally represent a small population of precise parameter settings, i.e., a precise view of cubes oriented at -45°, -30°, -15°, 0°, 15° and 30° about a vertical axis. Rather, it seems to me better to explicitly regard this as an example of crude parameter setting--the routine is sloppy enough that it will, for example, accept any cube which is roughly head on, say from -10° to 10° as satisfying a given parameter setting in the input-matching routine. More interestingly, we may imagine levels of precision, so that one range may be more precise than another. We thus require that the parameter sets

in both input-matching and action routines be partially ordered sets (posets, for short), where the relation $x \sqsubseteq x'$ is to be interpreted as x' is a refinement of x, or as x approximates x'. For example, if Y is a set of reachings, we might have

reaching to the left

$\sqsubseteq$ reaching about 60° to the left

$\sqsubseteq$ reaching 62° to the left while it is not true that "reaching 61° to the left

$\sqsubseteq$ reaching 62° to the left.

We also assume that each set has a minimal element 0

$0 \sqsubseteq y$ for all $y \in Y$

corresponding to 'no specification at all'--so that, in the reaching set, 0 just means 'reaching', without any specification as to the direction (and so is not to be confused with the very precise 'reaching 0° to the left', i.e., straight ahead).

For now, we shall assume that all schemas may continually monitor their input pathways (though different schemas have different input sets). In other words, the slide-file of the original metaphor becomes the total population of (relatively high-level) schemas of the present model; the slide-box of the original metaphor becomes the subpopulation of highly activated schemas of the present model. As in both Pandemonium and S-RETIC (recall Section 3), we let each schema (i.e., mode-element) continually receive input. However--unlike both Pandemonium and the original slide-box metaphor--we shall for now try to do without a central executive overseeing the activation of schemas and instead--in the spirit of Dev, Didday and S-RETIC--explore what can be achieved by the schemas themselves by virtue of their cooperation

and competition routines. My methodological point is that it is not helpful to make a priori assumptions (whether to fit our preconceptions about neural net structure or about the utility of LISP programming) when setting up a framework of this generality. When we actually look at restricted systems which must be implemented in a brain or on a computer, then we can be more specific about the sets of executive and book-keeping routines that seem necessary to augment the routines built into the schemas themselves.

With this, the time has come for a formal notion of scene to replace the "contents of the slide-box" of the old metaphor: A scene (A,e) is a set A of schemas together with a function

$$e: T \longrightarrow \prod_{a \in A} P_a$$

where T is a time interval, $P_a$ is the poset of parameter-settings of schema a, and for each $t \in T$,

$$e(t) = \{e(t)(a) \mid a \in A\}$$

is such that $e(t)(a) \in P_a$ is the parameter-setting of schema a at time t during scene (A,e).

Without being precise, it is part of the concept that the set A is relatively small--the scene (which is a potential new 'superscheme' in our learning theory for schemas [Arbib, 1976]) is the internal representation of a relatively homogeneous episode (i.e., one in which there are no 'dramatic' changes in the set of motors or their parameter settings). This does not require all schemas in A to be active throughout the scene-- $e(t)(a) = 0$ is permissible--but it does require that the variation of $e(t)(a)$ over time not change the structure of the situation too drastically (as would be the case if you were suddenly to perceive that a lion had entered your room).

Note that we have changed from the usual definition of scene--a two-dimensional visual input--to a definition that is system-dependent (different observers may activate different sets of slides in response to a given visual input) and extends over time. This is consistent with our general theme of action-oriented perception--a scene is to be a meaningful episode in the life of an organism interacting with a dynamic environment.

It is beyond the scope of this article to relate schemas to other approaches in the literature--but a few comments are in order. The problem of representation of knowledge has long been of great interest to psychologists. A classic study is Bartlett's [1932] analysis of "Remembering"; and we have gained much from Piaget's studies of schemata, especially his notions of assimilation and accommodation--see Furth [1969] for a review. These ideas have recently been born anew in approaches to artificial intelligence. Perhaps the best-known study is that of Winograd [1972, 1973], which works with a restricted 'blocks world'. Several other approaches to this general area are presented by Schank and Colby [1973]. Minsky [1975] has recently advanced his concept of 'frames' as a unification of these studies. At the level of vision and manipulation, these ideas seem bettered by our notion of a schema. However, the idea of a schema is not yet well tuned to the problems of linguistic and social interaction. Intriguingly, the notion of frame analysis proves not to be peculiar to artificial intelligence. Erving Goffman's [1974] "Frame Analysis: An Essay on the Organization of Experience" is a text in social psychology, and many of his examples are surprisingly reminiscent of Minsky's. A contribution to this area which seems to lie intermediate between the straight A.I. approach and Goffman's approach is that of Bruce and Schmidt [1974], which can be viewed as a sequel to Searle's [1965] study of 'speech acts'. Schank's work--and the related work of Abelson--are also interesting in this regard.

## 5. Competition and Cooperation between Schemas

Imagine that a segmentation program has divided a scene into regions such as those shown in Figure 9.

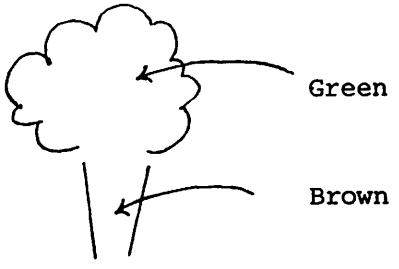



Figure 9

What is it?

With only this much information available, two quite different pairs of schemas may be activated to cover this input: in the first interpretation the schemas would represent green ice-cream and a brown ice-cream cone; in the second interpretation, the schemas would represent the foliage and trunk of a tree. There would be competition between the pairs, and cooperation between the schemas within each pair. Thus the system of interactions shown in Figure 10 would have two large attractors corresponding to the two natural interpretations, and very small attractors for the "unreal" pairings--though these could be forced by a trick photograph or a Magritte painting.

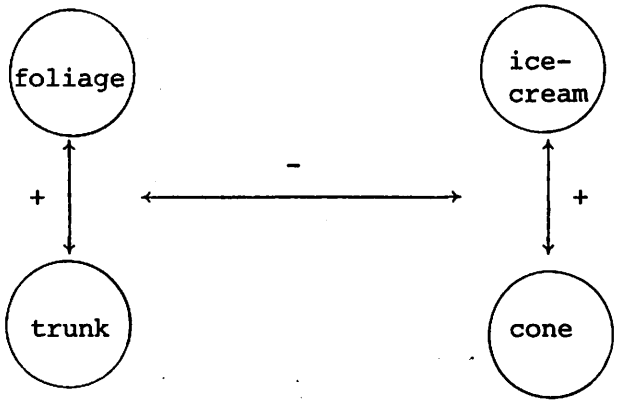



Figure 10

In case of blurred images, the input may start the system close to the boundary between the two attractors. Convergence may be slow, and even incorrect. Context can provide a mechanism to speed, and correct, convergence. Thus the 2 contexts in Figure 11a correspond to the 2 extra schemas in Figure 11b, and in each case we expect rapid convergence to the "right" interpretation.

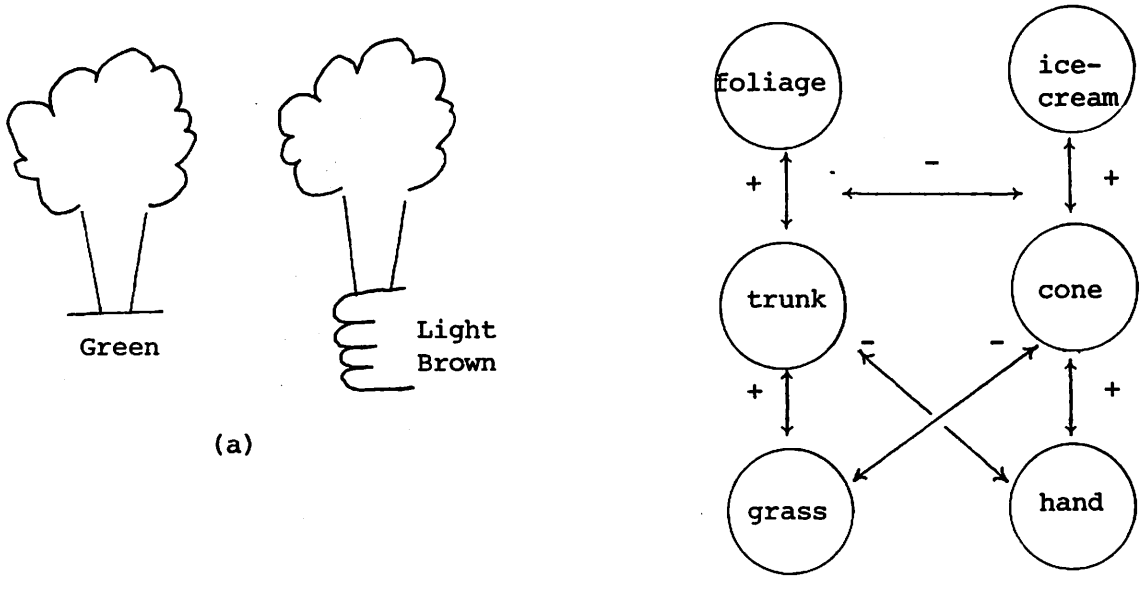



Figure 11

The effect of context

Thus an initial configuration in which the schemas for foliage, trunk, ice-cream and cone have comparable activity will rapidly converge to a state of high activity in foliage and trunk schemas and low activity in ice-cream and cone schemas if the grass schema is given a higher activity level than the hand schemas and vice versa. Incidentally, we may note that as well as slides for objects, we may also have more abstract schemas such as one for winter. Now at the change of seasons, the first fall of snow may be the signal for winter--so that we must posit the activity level of the snow-schema as providing excitatory input to the winter-schema. However, in the normal course of events, the organism knows that it is winter, and can use this contextual information (Figure 12) to favour the hypothesis that a white expanse is snow rather than burnished sand, say, or moonlit water. It is this type of reciprocal activation (whether we regard it as an additional input, or as the action of a cooperation routine) that gives the system of schemas its heterarchical character.

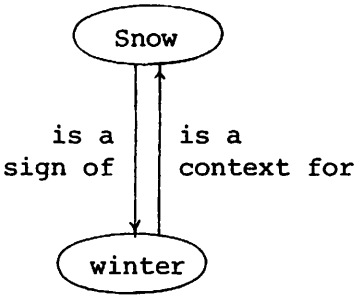



Figure 12

A Heterarchical Relationship.

[Strictly defined, a 'heterarchy' is a system of rule by alien leaders. But in AI, stimulated by McCulloch (1949), it now denotes a structure in which a subsystem A may dominate a subsystem B at one time, and yet be dominated by B at some other time.] This notion of schema competition and cooperation

may be seen to apply to the Dev model (Section 3) which we can recast in terms of 7 arrays of schemas (Figure 13).

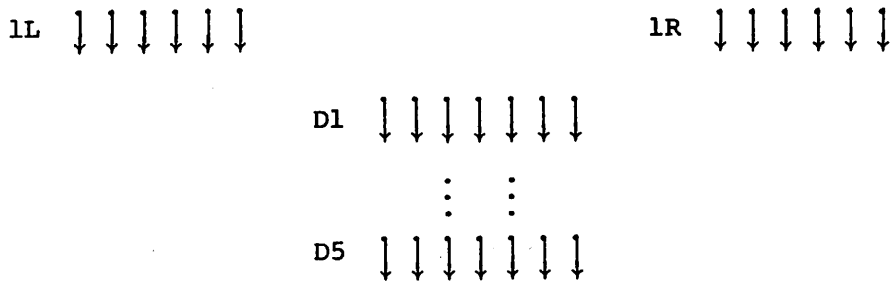



Figure 13

The Dev-model in schema format.

In this simple model, the schemas in layers 1L and 1R are simply ganglion cells of the retina, whose firing level is high to the extent that some feature is present in its visual field. The activity of a schema in layer Dk represents the presence of a feature with a given 3-dimensional location (as coded by the disparity between the schemas in 1L and 1R which activate it). But much of the activity here would be spurious if these schemas were driven only by the schemas of 1L and 1R--and so Dev postulates competition and cooperation routines (schemas in different D-layers compete; nearby schemas in a given D-layer cooperate) which yield segmentation of the visual field into relatively few regions in each of which the active D-schemas have the same feature.

Burt [1975] has modified the Dev model to support moving regions in response to moving inputs. In a more elaborate model one would then posit that--at this level--the visual field is not simply segmented into regions of relatively homogeneous schrmas with high activation, but that the cooperation routines have activated pointers to the activated schemas of the same region

(cf., the welds and co-moving sets of Arbib [1969]). Thus 'higher-level' schemas may determine that they are dealing with a region rather than an isolated feature.

Let us now leave this general discussion, and try to put competition and cooperation between schemas on a formal level. In one approach, due to Waltz [1975] and given a parallel algorithm by Rosenfeld, Hummel and Zucker [1975], the activation levels are 0 or 1, and one looks for a consistent labelling of the regions. In the second approach, also due to Rosenfeld et al., one imposes a nonlinear interation scheme on assignments of probabilities to the different labelling hypotheses for each region. The latter approach, as we shall see, is more satisfactory. Even more interestingly, however, is its strong formal resemblance to the S-RETIC scheme of section 3. Thus our study of competition and cooperation between schemas is brought firmly within our scheme of competition and cooperation in neural networks.

The Binary Model: We are given a set $A = \{a_1,\ldots,a_n\}$ of objects to be labelled, and a set $\Lambda = \{\lambda_1,\ldots,\lambda_m\}$ of labels. Let $\Lambda_1$ be the set of those labels in $\Lambda$ compatible with $a_i$; and let $\Lambda_{ij} \subset \Lambda_i \times \Lambda_j$ be the set of pairs of labels $(\lambda,\lambda')$ such that $\lambda$ may occur on $a_i$ when $\lambda'$ occurs on $a_j$. We set $\Lambda_{ii} = \Lambda_i \times \Lambda_i$.

[In our motivating example, we have 3 regions, with $\Lambda_1$ = {foliage, ice-cream}, $\Lambda_2$ = {trunk, cone}, $\Lambda_3$ = {grass, hand}, while $\Lambda_{12}$ = {(foliage, trunk), (ice-cream, cone)}, etc.]

A labelling $\underline{L} = (L_1,\ldots,L_n)$ assigns a set $L_1 \subset \Lambda_1$ of labels to each $a_i$. The labellings form a lattice under componentwise set-inclusion.

We say that a labelling is consistent if

$$(\{\lambda\} \times L_j) \cap \Lambda_{ij} \neq \emptyset \quad \text{for all} \quad \lambda \in L_i. \tag{1}$$

This is a local consistency condition. We call a labelling $\underline{\underline{L}}$ unambiguous if it is consistent and each $|L_i| = 1$. Note that a consistent labelling may not contain an unambiguous labelling. However, the task of the present model is to find such unambiguous labellings when they exist. We first note the obvious:

PROPOSITION The empty labelling is consistent. The union of any set of consistent labellings is again consistent. There is thus a greatest consistent labelling $\hat{\underline{\underline{L}}}$ (which may be null).

Returning to the criterion, (1), for consistency, let us say a label $\lambda$ in set $L_i$ is isolated in $\underline{\underline{L}}$ if $(\{\lambda\} \times L_j) \cap \Lambda_y = \emptyset$ for any j. It is clear that such a $\lambda$ cannot be part of a consistent sublabelling of $\underline{\underline{L}}$. This suggests that we operate on $\underline{\underline{L}}$ with $\Delta$ where $\Delta\underline{\underline{L}}$ is obtained from L by discarding from each $L_i$ all labels which are isolated in $\underline{\underline{L}}$. It is clear that $\Delta\underline{\underline{L}} = \underline{\underline{L}}$ iff $\underline{\underline{L}}$ is consistent. More interestingly, a standard fixed-point argument shows:

PROPOSITION For any $\underline{\underline{L}}$, let $\underline{\underline{L}}^{(\infty)}$ be the greatest consistent labelling contained in $\underline{\underline{L}}$. Then $\underline{\underline{L}}^{(\infty)}$ is the greatest fixed point of $\Delta$ contained in $\underline{\underline{L}}$, and, by the finiteness of $\Lambda$, we have that there exists an integer k such that

$$\underline{\underline{L}}^{(\infty)} = \Delta^k \underline{\underline{L}}.$$

To find an unambiguous labelling contained in $\underline{\underline{L}}$, we build a tree (in the style of Waltz [1975]) of labellings, with $\underline{\underline{L}}^{(\infty)}$ as the root. Then, having obtained a node with labelling $\overline{\underline{\underline{L}}}$, we construct its descendants by making all possible choices as follows:

Pick an i such that $\overline{L}_i$ is not a singleton. Pick $\lambda \in \overline{L}_i$. Set

$$\underline{\underline{L}}' = (\overline{L}_i, \ldots, \overline{L}_{i-1}, \{\lambda\}, \overline{L}_{i+1}, \ldots, \overline{L}_n).$$

If $\underline{\underline{L}}'^{(\infty)} \neq (\emptyset, \ldots, \emptyset)$, add it to the tree as a descendant of $\overline{\underline{\underline{L}}}$.

It is clear that all unambiguous labellings (if there are any) contained in $\underline{\underline{L}}$ will occur as (some of) the terminal nodes of the tree grown

in this way.

The main problem with this algorithm is that if any region has an empty label set $L_i$ in $\underline{L}$, then $\Delta\underline{L}$ will be the empty labelling. Clearly, a better algorithm would use the 'consensus' of other regions to add labels to $L_i$--as when we suddenly perceive the nature of an object purely on the basis of its context. The next algorithm, then, allows continuously varying weights to be assigned to each label for each region, and uses an operation which can increase, as well as decrease, those weights. As remarked before, the scheme is strongly reminiscent of S-RETIC, though here convergence is towards a labelling, rather than towards consensus on a single mode.

The Nonlinear Probabilistic Model: We again have a set $A = \{a_1,\ldots,a_n\}$ of regions, and a set $\Lambda = \{\lambda_1,\ldots,\lambda_m\}$ of labels. However, a labelling $\underline{p} = (p_1,\ldots,p_n)$ is now a sequence of probability vector $p_i: \Lambda \longrightarrow [0,1]$, with $p_i(\lambda)$ being the weight assigned by $\underline{p}$ to the hypothesis that $\lambda$ is the correct label for $a_i$.

We wish to design an operator F which--in the style suggested by our discussion of Figure 11--well on iterated application move $\underline{p}$ towards a 'correct' labelling. The key idea is that the probability $p_i(\lambda)$ of a given label for $a_i$ should be increased (respectively, decreased) by F if other objects that have high probability labels are highly compatible (respectively, incompatible) with $\lambda$ at $a_i$.

Thus, in the present model, the $\Lambda_{ij}$ are replaced by compatibility functions

$$r_{ij}: \Lambda \times \Lambda \longrightarrow [-1,1]$$

which function like correlations: if $\lambda'$ on $a_j$ frequently co-occurs with $\lambda$ on $a_i$, then $r_{ij}(\lambda,\lambda')$ is positive; if they rarely co-occur, $r_{ij}(\lambda,\lambda')$

is negative; and if their occurrences are independent, $r_{ij}(\lambda,\lambda') = 0$.

[It is clear that the memory structures required to produce the compatibility functions may be quite elaborate. Returning to our example of Figure 11, the system would have to use the observation that regions 1 and 2 were contiguous, with region 1 above region 2, to obtain estimates like

$r_{12}$(foliage, trunk) = 0.7

$r_{12}$(foliage, cone) = -0.8

$r_{12}$(ice-cream, cone) = 0.9, etc.

Incidentally, the very arbitrariness of these three numbers makes it clear that the F we are constructing must be structurally stable--small changes in the $r_{ij}$'s must rarely perturb convergence. Unfortunately, we do not yet have rigorous proofs of convergence--though computer simulations are encouraging--let alone structural stability.]

To satisfy the 'key idea', we define the 'change operator' $\mathcal{K}$ by

$$(\mathcal{K}\underline{p})_i(\lambda) = \sum_j d_{ij}\left[\sum_{\lambda'} r_{ij}(\lambda,\lambda')p_j(\lambda')\right]$$

where $d_{ij}$ is some choice of nonnegative coefficients with each $\sum_j d_{ij} = 1$. Each $\sum_{\lambda'} r_{ij}(\lambda,\lambda')p_j(\lambda')$ expresses the 'consensus' of the labelling of $a_j$ by $\underline{p}$ as to the direction in which $p_i(\lambda)$ should shift.

With this definition of $\mathcal{K}$, one possible choice for F is then

$$F\underline{p} = \mathcal{R}[\underline{p} + \mathcal{K}\underline{p}]$$

where the normalization operator $\mathcal{R}$ replaces each $q_i$ of a vector q a corresponding probability distribution $\mathcal{R}q_i$.

We imagine the following operation of this scheme:

(1) The segmentation routines divide the original routines into regions. Shape and texture descriptors are used to assign initial probabilities $p_i(\lambda)$ to appropriate labels $\lambda$ for each region $a_i$.

(2) Informaton such as relative position and the nature of the boundary would be used to generate the compatibility coefficients $r_{ij}$.

(3) F would be iterated a few dozen times, say, to provide enhanced probabilities. If the result is unambiguous, interaction with other systems--perhaps using higher-level context more subtle than that expressible in the $r_{ij}$--could be involved in disambiguation, with the possibility of reinitiating F using a new set of probabilities.

We have already noted the similarity of the nonlinear probabilistic model to the S-RETIC, but with the emphasis on 'proper labelling' rather than on 'mode consensus'. Re-viewing the Dev model of segmentation on prewired features in this light, we may note that it could be used for convergence to a sloping surface as well as for segmentation into regions of constant disparity. In fact, as Rosenfeld et al. [1975] note, their scheme can be used--as Dev has already done--to allow 'clusters' of low-level features having compatible labels to reinforce one another. For example, if the features are line orientations, we could use such a scheme to 'reinforce' those features which line up with their neighbors.

In conclusion, it seems that cooperative computation--a multi-level organization for problem-solving using many diverse, cooperating sources of knowledge, to use the title of the paper by Erman and Lesser [1975]--will provide the proper paradigm not only for the bottom-up and top-down approaches to brain theory (it seems to provide the right language for analyzing the neurological data of Luria [1973]), but also for artificial intelligence--a field which will provide many concepts for brain theory.

References

R. P. Abelson [1973] "The Structure of Belief Systems", in Schank and Colby, 287-339.

M. A. Arbib [1970] "Cognition--A Cybernetic Approach", Chapter 13 of Cognition: A Multiple View (Paul L. Garvin, ed.) Spartan Books, 331-348.

M. A. Arbib [1972] The Metaphorical Brain, Wiley-Interscience.

M. A. Arbib [1976] Brain Theory and Artificial Intelligence, Academic Press (To appear).

M. A. Arbib, C. C. Boylls and P. Dev [1974] "Neural Models of Spatial Perception and the Control of Movement", in Kybernetik und Bionik/Cybernetics and Bionics (W. D. Keidel, W. Handler, M. Spreng, eds.) R. Oldenbourg, 216-231.

H. B. Barlow, C. Blakemore and J. D. Pettigrew [1967] "The Neural Mechanism of Binocular Depth Discrimination", J. Physiol. 193, 327-42.

F. C. Bartlett [1932] Remembering, Cambridge University Press.

V. Braitenberg and N. Onesto [1960] The Cerebellar Cortex as a Timing Organ, Congress, Inst. Medicina Cibernetica, Naples, 239-255.

B. Bruce and C. F. Schmidt [1974] July 26-27 "Episode Understanding and Belief Guided Parsing", Association for Computational Linguistics Meeting, Amherst, MA.

P. Burt [1975] in press "Computer Simulation of a Dynamic Visual Perception Model", Int. J. Man-Machine Studies.

P. Dev [1975] in press "Segmentation Processes in Visual Perception: A Cooperative Neural Model", Int. J. Man-Machine Studies.

R. L. Didday [1970] "The Simulation & Modelling of Distributed Information Processing in the Frog Visual System", Ph.D. Thesis, Stanford Univ.

R. L. Didday and M. A. Arbib [1975] "Eye Movements and Visual Perception: a "Two Visual System" Model", Int. J. Man-Machine Studies (In press).

L. Erman and V. Lesser [1975] "A Multi-Level Organization for Problem-Solving Using Many, Diverse, Cooperating Sources of Knowledge", Proc. 4th Intl. Joint Conf. Artificial Intelligence.

H. G. Furth [1969] Piaget and Knowledge, Prentice-Hall.

E. Goffman [1974] Frame Analysis: An Essay on the Organization of Experience, Harper Colophon Books.

A. R. Hanson, E. M. Riseman and P. Nagin [1975] "Region Growing in Textured Outdoor Scenes", Proc. 3rd Milwaukee Symposium on Automatic Computation and Control, 407-417.

B. Julesz [1971] Foundations of Cyclopean Perception, University of Chicago Press.

W. L. Kilmer, W. S. McCulloch and J. Blum [1969] "A Model of the Vertebrate Central Command System", Int. J. Man-Machine Studies 1, 279-309.

J. Y. Lettvin et al. [1961] "Two Remarks on the Visual System of the Frog", in Sensory Communication (W. Rosenblith, ed.) M.I.T. Press.

A. R. Luria [1973] The Working Brain, Penguin Books.

W. S. McCulloch [1949] "A Heterarchy of Values Determined by the Topology of Nervous Nets", Bull. Math. Biophys. 11, 89-93.

M. L. Minsky [1975] "A Framework for Representing Knowledge", The Psychology of Computer Vision (P. H. Winston, ed.) McGraw-Hill.

F. S. Montalvo [1975] "Consensus vs. Competition in Neural Networks," Int. J. Man-Machine Studies 7, 333-346.

J. D. Pettigrew, T. Nikara, P.O. Bishop [1968] "Binocular Interaction on Single Units in Cat Striate Cortex", Exp. Brain Res. 6, 391-410.

W. H. Pitts and W. S. McCulloch [1947] "How We Know Universals: The Perception of Auditory and Visual Forms", Bull. Math. Biophys. 9, 127-147.

A. Rosenfeld, R. A. Hummel, S. W. Zucker [1975] May "Scene Labelling by Relaxation Operations", TR-379, Computer Science Center, University of Maryland, College Park.

R. C. Schank [1973] "Identification of Conceptualizations Underlying Natural Language", in Schank & Colby, 187-247.

R. C. Schank and K. M. Colby (eds.) [1973] Computer Models of Thought and Language, W. H. Freeman.

J. R. Searle [1965] "What is a Speech Act?", Philosophy in America (Max Black, ed.) Allen & Unwin, 221-239.

O. L. Selfridge [1959] "Pandemonium: A Paradigm for Learning", Mechanization of Thought Processes, London: H.M.S.P., 513-526.

H. R. Wilson and J. D. Cowan [1973] "A Mathematical Theory of the functional dynamics of cortical and thalamic nervous tissue", Kybernetik 13, 55-80.

T. Winograd [1972] Understanding Natural Language, Academic Press.

T. Winograd [1973] "A Procedural Model of Language Understanding", in Schank & Colby, 152-186.

# Parallelism, Slides, Schemas, and Frames

Michael A. Arbib
Computer and Information Science Department
Center for Systems Neuroscience
University of Massachusetts
Amherst, Massachusetts 01002

In BT (Brain Theory), we study nets of simultaneously active neurons, and of interacting brain regions. In AI (Artificial Intelligence), we must structure programs for a serial computer. However, the development of a serial algorithm for a function does not preclude the existence of a more efficient parallel algorithm. For example, when adding two numbers, the propagation of the carry bit seems to force seriality. However, a look-ahead adder (see Hill and Peterson [1973] for a textbook treatment) can be built which uses parallelism based on 'carry look-ahead' to reduce addition time from the order of n (the length of the numbers) to the order of log n which, in fact, is the best possible (cf. Winograd's [1965]). Our task here is to examine the ways in which behavior is best expressed in structure, and consider the extent to which we can expect parallelism in that structure. Clearly, the 'precedence relations' of the real world--you must walk to the door before you go through it, for example--impose a high-level seriality

on the flow of computation. However, within these high-level constraints, we shall see much room for parallel computation.

## 1. Parallelism

There is no question of the importance of parallelism in the early stages, at least, of visual processing. We see parallel extraction of 'bugness' in the frog retina, of contour and contrast information in mammals, and of other features in other animal visual systems. With their preprocessing cones, Riseman and Hanson [these proceedings] have demonstrated the usefulness of parallel computation in layered structures as a first stage in scene analysis.

Didday's model [1970] of the frog tectum gives a low-level example of parallel decision-making--a network of 'sameness' and 'newness' elements acts in parallel upon its input array to extract the strongest (with exceptions analogous to those seen in frog behavior). In scene analysis, however, the recognition of, and the choice between, local features is not sufficient. We must pass from local to semi-global features--as when the local features of a door-frame define the enclosed area as a space through which we can walk. Dev [1975] has studied parallel networks for segmenting a scene into regions in a 'semantics-free' way, by having elements responding to a given feature in nearby locations excite other nearby detectors of that feature and inhibit detectors of other features. This results in the partition of the overall scene into regions in each of which only one type of feature detector is dominantly active. Riseman and Hanson use iterated computation up and down their cones to grow lines or regions of given texture. Burt's [1975] studies of networks which represent and support the movement of objects may give us

clues as to how to use motion to aid region segmentation.

On the output side, we know that the brain uses activity in an array of motoneurons to control the populations of muscle fibres that constitute muscles. On the other hand, in AI the control of a stepping motor or rotary actuator does not seem to require inherent parallelism. However, there are other uses of parallelism. For example, Boylls [1975] modelled the cerebellum and its associated brain-stem nuclei as parameter-setting structures. We know that the basic algorithms for locomotion are in the spinal cord, but that the spinal animal does not 'shape' its steps properly. Stimulation of brainstem nuclei can increase muscular activity but--and this is the crucial point--Orlovsky [1972] found that in a walking animal, a muscle's activity is only increased during that phase of the step in which it should indeed be contracting. It is as if we have a motor control computer and a parameter-setting computer acting in parallel, but with the motor control computer only consulting the parameter setting when it is appropriate to do so.

Selfridge [1959] posited a character recognition system Pandemonium, which would behave as if there were a number of different 'demons' sampling the input. Each demon was an expert in recognizing a particular classification and would yell out the strength of its conviction. An executive demon would then decree that the input belonged to the class of whichever demon it heard yelling the loudest. On the other hand, Kilmer, McCulloch and Blum [1969], in modelling the reticular formation, posited a system without executive control. Rather, each of an array of modules sampled the input and made a preliminary decision to the relative weights of different modes as being appropriate to the overall commitment of the organism. The modules were then coupled in a back-and-forth fashion so that eventually a majority of the modules would

agree on the appropriate mode--at which stage the system would be committed to action. A reasonable analogy is a panel of physicians sharing symptoms and coming to a consensus about a diagnosis for a patient. (This suggests that social analogies may once again play an important role in brain theory.) Didday's [1970] model of the snapping behavior of a frog confronted with two flies, already mentioned in Section 1, posited a system of competitive interaction in the frog's tectum, which would lead in most cases to the suppression of all but one region of 'bugness' signalling, and result in the frog's snapping at one of the flies which caused the visual stimulation.

In attempting to place these studies in perspective, Montalvo [1975] observed that we could analyze all three models within a common framework, with the computational subsystems arrayed along two dimensions, one of competition and one of cooperation. In the Didday model, the cooperation dimension is 'degenerate' and the competition dimension is 'bug location'; in the Kilmer and McCulloch model competition is between 'modes' while cooperation is between 'modules'; and in the Dev module, competition is between 'disparities' and cooperation is along the 'space' dimension. For a careful treat the theme of competition and cooperation has thus emerged in three completely separate neural network models. It also plays a role when we look at the way in which an internal model of the world would operate.

Amongst the problems of a system receiving sensory input on the basis of which it must interact with the world are:

(i) Segmentation: To partition the input into 'segments' (not necessarily contiguous, nor confined to one modality) which define a single 'object' or other 'locus for possible interaction'. Dev [1975] provides a neural net model of segmentation processes in visual perception which shows how

cooperation (consensus mechanisms) and competition of feature detectors can form part of a very low-level input-matching process. In the preprocessing cones of Hanson and Riseman [1975], more subtle routines--operating in parallel up and down several layers of preprocessors--are being developed for segmentation on non-primitive features, such as texture. Burt [1975] has modified the Dev model to support moving regions in response to moving inputs.

(ii) Characterization: The 'segments' are to be characterized in terms of 'programs for possible interaction'. [Stages (i) and (ii) are by no means sequential--some success at characterization may well aid the aggregation of distinct regions into a single 'segment'.]

(iii) Relocation: As the system moves, or as objects move in its world, the system must be able to easily update the internal representation of its world to take account of these changes. An important part of the updating is that an object which is moving uniformly is expected to continue doing so--Burt [1975] has neural net models which can support such moving representations, although the representations do not yet have the complex slide structure which we shall outline below.

(iv) Tuneability: With further 'exploration', or with changing goals, the system can tune its internal representation, either by increasing the level of resolution, or by emphasizing those features most relevant to the current goal structure.

(v) Learning: The system should not only change its representation on the basis of new input, but should be able to change the way in which that representation is constructed.

Didday and Arbib [1973] have built upon Didday's model of the frog tectum to suggest that human eye movements are controlled in part by computation in the superior colliculus akin to that taking place in the frog tectum, but with 'bugness' being replaced by a combination of peripheral signals, hypothesis signals, and mismatch signals--with the latter two classes of signals descending from the cortex.

Rosenfeld et al.'s [1975] study of region labelling also falls within this general theory of competition and cooperation [Arbib, 1975a], and the notion of cooperative computation finds support in a number of neurological studies such as those of Geschwind [1965], Luria [1973] and Nauta [1971] (see Arbib, 1975b).

## 2. Slides and Schemas

The basic slide-box metaphor (Arbib [1972, p. 92]) was intended as an antidote to a simple pattern recognition system in which the visual input pattern was to be classified as belonging to one of a small number of classes. Rather, the input pattern was to be analyzed as an array of familiar 'objects'--whether the object was a single object, such as a tree, or a composite object such as a row-of-trees--retrieving slides from a file, and arraying them in a slide-box to represent the current scene.

A critical notion was that covering a portion of sensory input was to give access to appropriate programs for action--though, since there are many objects, only some of the programs can 'take control' at any time, and some of them would be planning programs rather than programs for overt action. Thus some mechanisms must restrict which programs amongst this redundancy of potential command (to use McCulloch's phrase) will actually be implemented.

In an AI context, Minsky [1975] has developed a concept of frame which in some ways overlaps the above concept of slide, though with far more emphasis on linguistic and sociological aspects. We shall discuss frames in more detail in Section 3. (For more on computer understanding of language, see Schank and Colby [1973].) Intriguingly, the sociologist Erving Goffman 1974 has independently coined the term 'frame analysis' for analyzing the organization of experience--and his analysis has many points of overlap with Minsky's. Relevant ideas also occur in the approach to scene analysis espoused by Hanson and Riseman [1975].

In the rest of this section we outline an updated slide-box model for the organization of action-oriented memory for a perceiving system. A fuller account appears in Arbib [1975a,b]. To avoid the overly pictorial connotations of the term "slide," I have adopted Piaget's term "schema," since much of the flavor of the tie-up between input-matching routines and action routines is contained in Piaget's notion of a schema (see Furth [1969] for an exposition). My concept is more formal and will hopefully (!) be shown to encompass all the more valuable aspects of Piaget's notion. I thus regard a schema henceforth as an array of programs to analyze a segment of the input to determine a possible course of action. As such, a schema must be locatable, tuneable, and linkable with other schemas. Thus, we have the following three components of a schema:

(i) Input-Matching Routines: A routine which succeeds to the extent that it covers a spatial region of multi-modal input (so that a cat schema can cover a furry region which emits meows, but not one that emits barks). This can be biased by non-sensory context inputs. The input-matching routines may include calls for confirming information (as in the eye-movement calls of Didday and Arbib [1975]--see also Szentágothai and Arbib [1974, pp. 335-339]). The level of success of these input-matching routines may be regarded as an 'activation' level of the schema, which increases as location and other parameters of the schema

are adjusted to better fit the covered region. However, to complicate the story, the resolution level (i.e., the precision of this parameter match) required for activation to saturate may well depend on goal settings, or other non-sensory input to the schema. (In the vision routines of Hanson and Riseman [1975], the sensory input is purely visual--though contextual input is also used--and 'success' simply enables the assignment of a name to a region.)

(ii) Action Routines: The success of the input-matching routines in raising the activation level of a schema signals that certain actions have become appropriate for the system. Programs for some of these actions then form part of the schema. A crucial integrative property of schemas is that increasing accuracy of parameter adjustment by the input-matching routines automatically adjusts parameters in the action routines in such a way that the action becomes more appropriate for the current environment and goal structure (if the schema has been properly 'evolved').

(iii) Competition and Cooperation Routines: To date, we have talked of a schema as acting in isolation, attempting to raise its activation level by proper matching of input. But now we must realize that schemas are interconnected. The operation of competition and cooperation routines helps determine which population of parameter-adjusted highly active schemas will constitute the current model of the environment. There still remains the problem of determining which of the action routines of these schemas are to operate--and another network of competition and cooperation routines will be

involved in determining a compatible set of actions. (Since at any time, the organism is engaged in some activity, even if that be resting. Thus it is not so much a matter of choosing a course of action as it is of determining whether the time has come to change the course of action being pursued. The completion of an action may remove it from the competition. More interestingly, the execution of an action may provide new sensory input which de-activates the schema (or drastically changes the parameter settings) which enable the action--as when we bite into what appears to be a piece of fruit only to discover that it is made of wax).)

Notice that all these routines provide the semantics of a schema--what an object means to us comprises our knowledge of what we can do with the object and what relations it has with other objects, to the extent that our input-matching routines can capture the effects of these actions and relationships. In any case, we've come a long way from the original notion of a slide as being simply a colored transparency that approximates a region of the visual field.

For now, we shall assume that all schemas may continually monitor their input pathways (though different schemas have different input sets). In other words, the slide-file of the original metaphor becomes the total population of (relatively high-level) schemas of the present model; the slide-box of the original metaphor becomes the subpopulation of highly activated schemas of the present model. As in both Pandemonium and RETIC, we let each schemas (i.e.,mode-element)continually receive input. However--unlike both Pandemonium and the original slide-box metaphor--we shall for now try to do without a central executive overseeing the activation of schemas, and instead--

in the spirit of RETIC--explore what can be achieved by the schemas themselves by virtue of their cooperation and competition routines. My methodological point is that it is not helpful to make a priori assumptions (whether to fit our preconceptions about neural net structure or about the utility of LISP programming) when setting up a framework of this generality. When we actually look at restricted systems which must be implemented in a brain or on a computer, then we can be more specific about the sets of executive and book-keeping routines that seem necessary to augment the routines built into the schemas themselves.

In addition to schemas for objects, we may also have more abstract schemas (cf. the Hanson-Riseman context routines) such as one for 'winter'. Now at the change of seasons, the first fall of snow may be the signal for winter--so that we must posit the activity level of the 'snow-schema' providing an input to the 'winter-schema'. However, in the normal course of events, the organism knows that it is winter, and can use this contextual information to favor the hypothesis that a white expanse is snow rather than burnished sand, say, or moonlit water. It is this type of reciprocal activation (whether we regard it as an additional input, or as the action of a cooperation routine) that gives the system of schemas its heterarchical character. [Strictly defined, a 'heterarchy' is a system of rule by alien leaders. But in AI, stimulated by McCulloch (1949), it now denotes a structure in which a subsystem A may dominate a subsystem B at one time, and yet be dominated by B at some other time.] To the extent that the activation of a small population of schemas covers the activity of the feature-region schemas, to that extent can we say that the organism has perceived the scene.

## 3. Frames and Schemas

Given the complexity of physiological mechanisms that animals have evolved, we should expect the brain to be similarly sophisticated. Minsky [1975] thus posits a host of special-purpose mechanisms rather than a single simple mechanism. He introduces frames as his candidate for the unit underlying the effectiveness of common sense thought.

There seem to be three main reactions to Minsky's "Frames" paper:

(a) The "what a revelation" reaction of neophytes who had never before realized the importance of an internal representation of the world. Having confined their reading to a few recent theses and papers in AI, they were unaware of such contributions (to give but a limited sample) as Bartlett [1932], Craik [1943], Gregory [1969], MacKay [1955, 1963], Piaget [1954], Minsky [1961, 1965], and Young [1964].

(b) The "we've seen it all before" reaction. This comes in two flavours.

  (i) Some AI experts, having developed their own formalism for handling internal representations, dismiss Minsky's frames as a vague equivalent to their precise formulation.

  (ii) Other readers, familiar with the literature in (a), feel that Minsky was too cavalier in his brief reference to these earlier works, and object that many aspects of frames are ideas about internal representations with which they are long familiar.

(c) The "mature acceptance" reaction of workers in AI who have felt the need for a more general framework for the discussion of internal representations, and feel that recasting their work in the language of frames is a reasonable price to pay in moving toward such generality.

Having invested a moderate amount of effort in the slide-box metaphor and its extension to theory of schemas of Section 2, I must confess that my initial reaction was (b.ii). It is my feeling that Minsky's treatment of frames for scene analysis is inferior to mine in the sense that it seems too computer

oriented rather than general enough to respond satisfactorily to the needs of brain theory. However, I recognize that Minsky's discussion of frames for language, understanding, and scenarios builds on recent work in computer understanding of natural language to handle aspects of internal representations for which my slides and schemata are too concrete. To this extent, I sympathize with reaction (c). The resolution of these apparently incompatible reactions is to distinguish more carefully than Minsky does between scene analysis and language understanding and discuss the extent to which they demand different styles of internal representation, the former being more slide/schema-like, the latter being more frame-like. The rest of this section discusses frame and schema for scene analysis; while Section 5 discusses pressures which require "protolinguistic" extensions to be made to the concept of a schema.

A frame includes information about

How to use the frame

What one can expect to happen next

What to do if these expectations are not confirmed.

The top level represents things that are always true about the situation. Lower levels have <u>terminals</u> (slots) which are usually filled by 'subframes' which must meet certain conditions assigned at the terminal. In fact, a set of terminals may impose relations on their mutual assignments.

To handle the dynamics of a changing world Minsky posits a <u>system</u> of frames, transformations between which mirror the effects of important actions or changes in the world. Thus: a frame-system $\doteqdot$ a schema, and a transformation $\doteqdot$ updating schema parameters.

The theory of frames must handle change of percepts in the face of inconsistencies, errors, or new evidence; must explain imagery (cf., Bartlett [1932]) and must show how to exploit expectations. Minsky suggests that, to handle expectations, a frame's terminals are normally filled with default assignments--i.e., one cannot think of a ball without thinking of, say, a soccer ball of given size and

colour. However Arbib [1975a] instead argues for a poset of schema-parameter specifications--and I would suggest that 'default' assignments may be very 'blurry' elements of the poset indeed, so that a ball may be little more specified than as requiring 'two hands to hold it', rather than being 'one-hand-holdable'. This is far from the level of precision that Minsky seems to suggest.

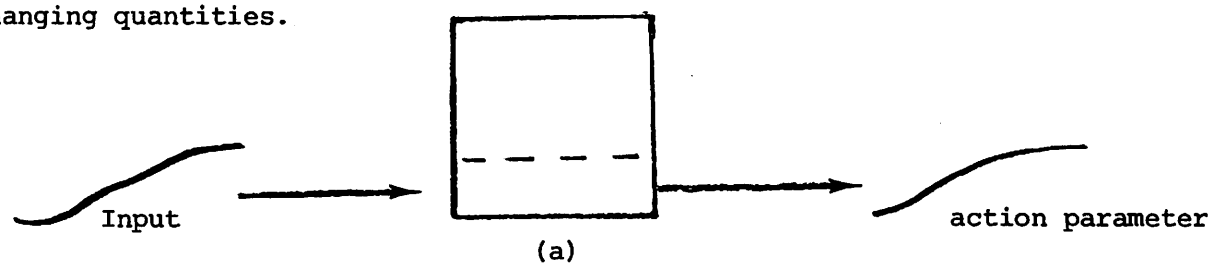
However if Minsky ascribes precision to each frame, he is insistent that the frame-system is far less precise than people believe their perception to be, citing the inability of all but the best draftsman to precisely render a variety of perspectives of a given object, and thus suggesting that a frame system comprises very few frames indeed. Minsky does not believe that our image changes as fast as does the scene. Rather, he posits that the illusion of continuity is due to the persistence of assignments to terminals common to different view frames--so that 'continuity' depends on the confirmation of expectations which in turn depends on rapid access to remembered knowledge about the visual world.

This analysis via discrete frames may be misguided. Stressing again our action-oriented view of perception, it is not the natural task of the system to draw pictures or to judge the accuracy of a drawing--though draftsmen can indeed master these skills. Rather, the task of the input matching routines is to activate schemas in a way which sets the parameters of action routines appropriate to the sensory input. Thus our shortcomings as draftsmen in no way imply a limitation of the accuracy of parameter-setting when we interact with real objects. We need not postulate an ability of the schemas to maintain very precise metrical relationships, for once foveal scanning has activated a schema, peripheral input--even though inadequate for object recognition--will suffice to maintain the appropriate place information when the object is no longer fixated.

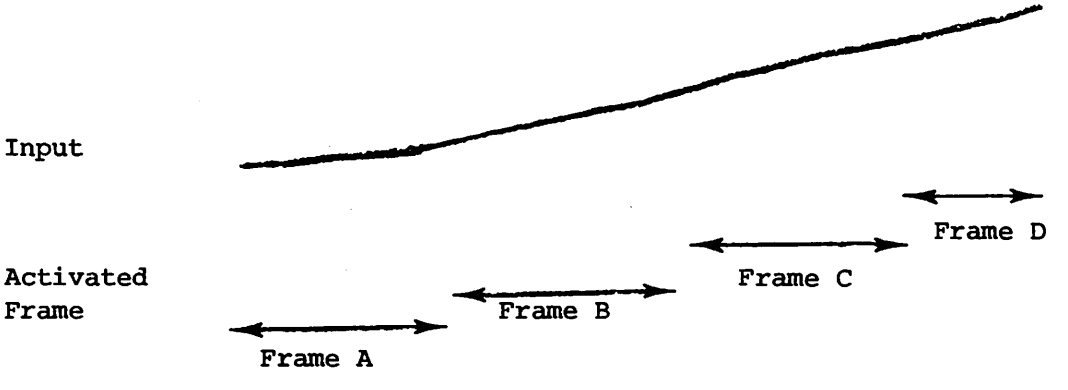
Our viewpoint, then, is that an organism interacting with its world does not need a complete representation, but rather one that is easily updated as action progresses. [One of the greatest problems for the Shaky robot project was the lack of continuous visual input. In the same way, it is far easier to

walk through a crowded room with our eyes open than it is to memorize the scene in sufficient detail to allow us to close our eyes and then walk to the door without bumping into anyone.] It is the range of actions in which the system will take part that will determine the appropriate level of detail--a map may have the metric relations all wrong so long as it reminds us to take the correct turning when we traverse the actual terrain--and so we see the effect of 'goal-setting schemas' upon the level of parameter-matching that will let a slide be sufficiently activated for its action-routines to be candidates for implementation.

In Section 1 we mentioned the work of Orlovsky as one dramatic instance of neural parameter-setting. It is thus natural for the brain theorist to posit a tuneable system--once an object is identified, we simply update the parameters of the schema which represents it. By contrast (Figure 1) a frame represents an aspect of an object--with continuous changes in the input triggering discrete charges of frame. While this interchange of discrete structures may have some appeal for computer implementation, it appears unnatural for the brain, where varying frequencies of neural firing seem so well suited to represent continually changing quantities.




(a)




(b)

Figure 1
Schemas (a) vs. Frames (b)

This leads us to the second crucial concept in the schema notion--it stresses the generative nature of the internal representations, which are built up as 'collages' of schemas. Minsky's framework seems to lack this concept--he talks of having one frame accessed at a time. For example, he posits one frame (in the form of a face-relation graph) for each view of a cube, and a different frame to represent the manipulative features of a cube. Once a frame is proposed, a matching process tries to assign values to its terminals--an information retrieval network provides a replacement frame if matching fails. More specifically, Minsky posits the following process:

(1) Once a frame is evoked, it directs a test of its appropriateness, using the current goal list to decide which terminals must match reality.

(2) It requests information to assign values to those terminals. (Failure may cue evocation of an alternative frame.)

(3) Informed of a transformation (e.g., an impending motion) it would transfer control to the appropriate frame of the frame system.

We thus see an emphasis on frames as 'the complete internal representations of the current sensory input per se' rather than as 'a component of a representation which prepares the organism for interaction'. The emphasis is on a serial process of frame selection rather than the parallel process of competition and cooperation posited in the schema model; and little distinction is made between changing a parameter and changing a hypothesis.

We may concede that a room provides a frame within which the recognition of doors and windows and their relationships is simplified. However, in many cases, the framework is unimportant, and we directly recognize a number of objects and come to grasp the situation in terms of their directly perceived relationships. When we find an elephant sitting on our best chair, we realize what is happening in spite of the framework of our expectations. Our shock may be a measure of this discrepancy; but our understanding is a strong argument for the "college" approach as an important component of internal representations:

schemas trigger frames order schemas... It is in this more liberal sense of a frame as a surrounding that the concept seems most useful; with that which fits in the frame, the schema, have somewhat different properties. To the extent that we represent changes in objects, parameter-setting in schemas seems appropriate. But the decomposition of the world into objects imposes a discrete structuring onto a continuous world, and the relationships between objects require a frame-like representation closer to linguistics than to control theory. Perhaps a reasonable subgoal for a theory of representations is to analyze the extent to which there is a genuine distinction here and to what extent we are looking at poles of a continuum.

4. Development

Turning to developmental questions, Minsky suggests that we compare Piaget's concrete operations to the transformations between frames of a system (the tuning of parameters in a given schema); while the formal stage might be characterized in terms of the ability to reason about, rather than simply to work with, those transformations. In computer terms, we might speak of the system developing a facility for writing 'commentaries' on its programs, or for reading its own programs. One might imagine that an 'abstract' of each schema comes to be developed with the schema itself, and that these are then available to 'higher-level' schemas. Is the language of frame-systems or schemas appropriate to the study of 'representations of representations'? I suspect that the answer is 'Yes'.

We have stressed the idea of a heterarchy of active schemas, all the way from 'line detectors' to abstract concepts like 'winter'. The input to a 'snow' schema is as much the activity of a context-schema like 'winter' as it is the output of some texture slide. Thus we begin to dissolve a strict interpretation of input-matching routines as matching sensory input, or of action routines as controlling motor output. Instead, we have a general system which may be involved in monitoring some schemas to better adjust the activity of others, rather than in sensorimotor correlation. But this is still consistent with our general definition of a schema. Thus schemas at one level can form "inter-schema operators" for schemas at another level. Such interschema operators would seem crucial in any theory of learning based on explicit hypothesis formation rather than mere synaptic adjustment.

Another Piagetian problem is addressed by Minsky in discussing occlusion. There might be ad-hoc information about occlusion represented in a frame-system--so that we might have frames for each view of a chair being progressively occluced as it is slid under a table. Such frames probably do 'exist', but--more radically--

one might posit a GLOBAL OCCLUSION SYSTEM which makes all perspective frames subsidiary to a central, common, space-frame system. The terminals of that subsystem would correspond to cells of a gross subjective space, whose transformations represent, once and for all, facts about which cells occlude others from different viewpoints. This certainly seems plausible for humans, and the work of Piaget and Inhelder suggest that complete coordination structures of this sort are not available to most children until they are at least ten years old. However, it would be naive to over-estimate the accuracy of this space-fram even in adults. Nepalese villagers never identify the two faces of a mountain if the faces can only be viewed from two different valleys separated by such ragged mountains that one must travel 100 miles to go from one vantage-point to the next; and many city-dwellers may drive a mile to get from one building to another down twisting one-way streets without realizing that they are within easy walking-distance of one another.

One of the really difficult problems of AI is to model the development of the space concept without building mechanisms into the original system which trivialize the whole problem. Consider, for example, how Ernst and Newell [1969] trivialized the monkey-and-banana problem when they implemented it in GPS. The monkey discovers that by moving a box and standing on that box, it can reach a banana suspended from the ceiling, and otherwise out of reach. But by providing their GPS simulation with 'vertical height' as an explicit difference, and by making 'climbing on the box' the only operator available to reduce that difference, Ernst and Newell excluded the only process that was genuinely of interest--the monkey's discovery that it could use the box as a tool. The question (being actively tackled by my colleague William Kilmer) is how to give the system so little information that the discovery is 'impressive', and yet enough information to enable the discovery to be made. The idea is to give the monkey a body-centered frame of reference,

plus the ability to run, to bump into things, to reach and grasp, and to climb. It can then discover, as a result of several 'crashes', that boxes are moveable, and learn--through play--how to move them in a deliberate way. Again, it may learn that climbing on boxes gets it further from the ground. It is a challenge to our ability not to build too much in to make it a separate discovery--but one not so surprising in a body-centered, as distinct from Euclidian, framework--that this climbing also gets it closer to the ceiling. Then, and only then, is it ready to solve the problem of reaching the bananas hanging from the ceiling. Note here the importance of play in providing mechanisms for later problem-solving.

## 5. More on Parallelism

Didday and Arbib [1975] studied eye movements and visual perception with a hybrid model in which a 'slide-box' cortex interacted with a midbrain system adapted from Didday's neural net model of the frog tectum. The constraints imposed on cortex by interfacing it with a midbrain system so structured will suggest ways to move towards a more neural model of cortex.

We perceive a scene via a series of visual fixations, required to bring successive regions before the fovea; although the above model suggests that the computation required to determine the eye movement and process the input is parallel. Is there any reason why the whole retina does not have foveal acuity, allowing the whole process of analyzing a scene to be accomplished in parallel upon a single fixation? The frog's 'bug-detectors' operate in parallel without eye movements. But a 'SUPERFROG' should not simply snap at flies, but should also learn about new objects in its world. If the whole recognition machinery were iterated over the whole visual field, there would be the problem of communicating information learnt about an object which has appeared in one part of the visual field to the machinery which would handle its appearance in each other region of the visual field. It may well be [a careful mathematical analysis is called for] that it is computationally effective to use, e.g., eye movements to route visual input to a standard processor then it is to route learned information to a host of parallel processors for the recognition of a given class of objects.

Seriality, then, is imposed on visual perception by the sequence of eye movements in visual perception; and we may note, too, the seriality of speech, as if each word were trying to direct a fixation of the attention of the listener. Arbib [1975a] discusses the notion of building a 'superschema' from a

repeated scene, thus providing larger units of representational activity. This accords well with Minsky's view that rapid selection of large substructures--to provide the context in which selection of 'subframes' takes place--will speed perception and thought. Though we disagree with Minsky's insistence on purely serial processing, the time has come to face up to the fact that there must be limits to the parallelism which a slide-box or collage of schema can handle.

We have talked of a schema for each concept, be it 'tree','winter', 'differential calculus' or 'ontological commitment'; and we've offered the set of currently active schemas, together with their parameter settings, as providing the internal representation of the current state--both internal and external--of the organism's perceived world. But how do we handle multiple objects? Do we imagine that we have several copies of each schema, and that the competition routines linking them are sufficiently strong that only n can be activated when n of the objects that schema represents are in the environment? Is it, then, that for each object, there is an upper bound on the number of instances we can apprehend--17 trees, and no more?! Certainly, the discussion of 'superfrog' makes this option unappealing--with multiple tree schemas, how do we share the adaptive changes in one with the other 16?

Perhaps--and I do not yet know how to phrase this in neural terms, but must use a simple-minded computerese--we should imagine a single copy of each schema, but posit that it can accomodate several pointers, with appropriate settings of location and other parameter settings for each object which provides the 'source' of such a pointer. Calling on the folklore of anthropology--stories of primitive tribes that count 'one, two, three, many'--it seems reasonable to posit an upper bound of three, say, to the number of pointers which can bear detailed parametric information. After this, a 'lumped' form of description--

a 'row-of-trees', say--may be the most explicit representation one can handle without linguistic intervention. Thus a human can count 17 trees, and remember that there are indeed seventeen--but this is a more abstract form of representation than the process of parameter tuning by input matching routines. [It is intriguing to speculate on the extent to which these two types of parameter settings--linguistic and non-linguistic--may be localized in the left and right hemispheres of humans; and to explore the role of the corpus callosum in integrating these two types of information. It may be--returning to our discussion of Piaget--that the distinction between the hemipsheres is related to the distinction between formal and concrete operations. However, this is probably too crude a division of labour.]

In any case, we see that schema must be able to 'quell' several regions, rather than simply one; and that a significant step in evolution--to avoid undue demands on parallelism--was the ability to move from 'quelling' by precise parameter adjustment by the input-matching routines to 'quelling' by a more abstract, proto-linguistic, representation which could, for example, simply note the number and approximate disposition of an array of similar objects. [Incidentally, I would suggest that this multitude of simultaneous activity is what makes perception 'richer' than imagining--we not only have 'tree' schemas active during perception, but a rich array of texture and other 'low-level' schemas, too. Note, too, that dreaming is a natural facet of the schema model: schemas can activate one another in complex activity patterns even with the reduced sensory input that characterizes sleep.]

The transition to proto-linguistic parameter is one way of overloading the capacity of any one schema. Another approach is to make 'copies' of the schema

tuned to different instances of a given type--as when we differentiate the schema for 'man' into one for each of the men with which we are at all acquainted, from a very sketchy schema for a public figure seen occasionally in the newspaper, to a schema, far richer than the generic schema, for a very close friend.

What does this say for the concept of identity? At one level, we may say an object, or a low-level pattern of schema activity, is more-or-less identical to another to the extent that--in a given context--it yields the same pattern of action or high-level activity. However, it is one thing for the organism to behave as if the two are identical; it another thing for it to be aware of this identity. Presumably, this 'awareness' requires the 'schema-abstracts' posited in the discussion of Piaget's formal operations, together with mechanisms to compared 'abstracts' activated by the two patterns.

In his article, 'The Architecture of Complexity', H. A. Simon argues for hierarchical structuring of complex systems, suggesting that evolution can more effectively act upon a system made up of functionally well-defined subsystems. Minsky (at this N.Y.U. symposium on 'Parallel Processing in AI') argued similarly, going from the need to debug knowledge systems (cf. Winston's program for learning structural descriptions by debugging a preliminary description by using examples and near-misses) to the need for structured programs. However, he seems mistaken when he suggests (admittedly in the role of 'Devil's Advocate') that such a structured program must be serial--any more than the evolution of organs should require heart, lungs, and liver to be time-shared!

Between the admittedly parallel input and output structures lies the

region of 'cognitive computation', and Minsky claims that this is inherently serial. In fact, we know that different regions of the brain communicate during cognition--one can get auditory tuning curves from visual cortex neurons--as if each region were trying to model the world on the basis of its own primary data, and yet keep the model consistent with information about the activity of other model builders (so that the 'internal model' in our heads is not a unitary construct, but is a population of models, agreeing in crude outline, but differing in type and depth of detail). However, it is striking that while this array of parallel subsystems lets us, for example, recognize with alacrity a sought-for object in a complex scene, our billions of neurons may take several seconds to add a pair of 4-digit numbers. This suggests that the natural parallelism expressed in our brains by the multiplicity of anatomically distinct regions may have evolved to suit us for a primitive hunting existence, but be little adapted for the linguistic and cultural 'computations' which mankind has evolved since our brains achieved their present form. This may impose a semblance of seriality on many such 'computations'. However--recall the look-ahead adder--this does not preclude the incorporation of far greater parallelism in a computational structure specifically designed for socio/linguistic behavior.

In summary, our concern in both AI and BT is with the mediation of complex behavior by appropriate structures. In each case, questions of efficiency, evolution, learning, and 'debuggability' will enter, and it can be expected that the temporally serial execution of a variety of processes operating in parallel will provide the proper setting for their analysis.

References

M. A. Arbib [1972] The Metaphorical Brain, Wiley-Interscience.

M. A. Arbib [1975a] "Segmentation, Schemas, and Cooperative Computation", MAA Studies in Mathematics, Biomathematics (S. Levin, Ed.).

M. A. Arbib [1975b] "Artificial Intelligence & Brain Theory: Unities & Diversities", Ann. Biomed. Eng.

F. C. Bartlett [1932] Remembering, Cambridge University Press.

C. C. Boylls [1975] "The Function of the Cerebellum and its Related Nuclei as Embedded in a General Paradigm for Motor Control", Technical Report, Computer and Information Science, University of Massachusetts at Amherst.

P. Burt [1975] "Computer Simulation of a Dynamic Visual Perception Model", Int. J. Man-Machine Studies, in press.

K. J. W. Craik [1943] The Nature of Explanation, Cambridge University Press.

P. Dev [1975] "Segmentation Processes in Visual Perception: A Cooperative Neural Model", Int. J. Man-Machine Studies, in press.

R. L. Didday [1970] "The Simulation & Modelling of Distributed Information Processing in the Frog Visual System", Ph.D. Thesis, Stanford Univ.

R. L. Didday and M. A. Arbib [1975] "Eye Movements and Visual Perception: a "Two Visual System" Model", Int. J. Man-Machine Studies, in press.

G. W. Ernst and A. W. Newell [1969] GPS: A Case Study in Generality and Problem Solving, Academic Press.

H. G. Furth [1969] Piaget and Knowledge, Prentice-Hall.

N. Geschwind [1965] "Disconnexion Syndromes in Animals and Man", Brain 88: Part I, 237-294, Part II, 585-644.

E. Goffman [1974] Frame Analysis: An Essay on the Organization of Experience, Harper Colophon Books.

R. L. Gregory [1969] On How so Little Information Controls so Much Behavior, in Towards a Theoretical Biology 2 Sketches, (C. H. Waddington, Ed.), Edinburgh University Press.

A. L. Hanson and E. M. Riseman [1975] "The Design of a Semantically-Directed Vision Processor", COINS Technical Report 75C-1, University of Massachusetts at Amherst.

F. J. Hill & G. R. Peterson [1973] Digital Systems: Hardware Organization and Design, Wiley.

W. L. Kilmer, W. S. McCulloch and J. Blum [1969] "A Model of the Vertebrate Central Command System", Int. J. Man-Machine Studies 1: 279-309.

A. R. Luria [1973] The Working Brain, Penguin Books.

D. M. MacKay [1955] The Epistemological Problem for Automaton, in Automata Studies, (C. E. Shannon and J. McCarthy, Ed.), Princeton University Press.

D. M. MacKay [1963] Internal Representation of the External World, AGARD Symposium on Natural and Artificial Logic Processors, Athens, Mimeographed, 14 pages.

M. L. Minsky [1961] Steps Towards Artificial Intelligence, Proc. IRE 49, 8-30.

M. L. Minsky [1965] Matter, Mind and Models, Information Processing 1965, Proc. IFIP Congress, 1, 45-49.

M. L. Minsky [1975] "A Framework for Representing Knowledge:, The Psychology of Computer Vision, McGraw-Hill (P. H. Winston, Ed.) 211-277.

F. S. Montalvo [1975] "Consensus vs. Competition in Neural Networks," Int. J. Man-Machine Studies, 7, 333-346

W. J. H. Nauta [1971] "The Problem of the Frontal Lobe: A Reinterpretation", J. Psychiat. Res., 8: 167-187.

G. N. Orlovsky [1972] "The Effect of Different Descending Systems on Flexor and Extensor Activity During Locomotion", Brain Research 40: 359-372.

J. Piaget [1954]"The Construction of Reality in the Child,"Basic Books.

R. C. Schank and K. M. Colby (Editors) [1973] Computer Models of Thought and Language, W. H. Freeman.

O. L. Selfridge [1959] "Pandemonium: A Paradigm for Learning", Mechanization of Thought Processes, London: H.M.S.P., 513-526.

J. Szentágothai and M. A. Arbib [1974] Oct. Conceptual Models of Neural Organization, Neurosciences Research Program Bulletin 12: No. 3.

S. Winograd [1965] "On the Time Required to Perform Addition", J. Assoc. Comp. Mach. 12: 235-243.

J. Z. Young [1964] "A Model of the Brain", Oxford University Press.