

ARTIFICIAL INTELLIGENCE AND BRAIN THEORY:

UNITIES AND DIVERSITIES*

Michael A. Arbib
Computer & Information Science Department
Center for Systems Neuroscience

COINS Technical Report 75C-8

(September 1975)

* The material in this paper will be expanded to book length in the author's Brain Theory and Artificial Intelligence (Academic Press, 1976), which will contain fuller details of many of the models sketched herein. The research reported in this paper was supported in part by NIH 5R01 NS09755-06 COM.

ABSTRACT

This review article synthesizes studies of artificial intelligence (AI) and brain theory (BT). In the control of movement, AI offers insight into overall planning of behavior; while control theory enables BT to model feedback and feedforward adjustments by the spinal cord, brainstem and cerebellum. We stress action-oriented perception--analyzing perception in terms of preparation for interaction with the world, and offer a new concept of a schema as the internal representation of an 'object' in the sense of a domain of interaction. A schema comprises input-matching routines, action routines, and competition and cooperation routines. The internal representation of the world is then given by a 'collage' of tuned and activated schemas. Segmentation of input and region labelling are offered as two mechanisms in the activation of a suitable 'collage'.

We see a number of studies which offer hope of a unified theory of competition and cooperation within a single subsystem: brain theory models of the reticular formation, of the frog midbrain visual system, and of segmentation on prewired features; and AI models of segmentation on ad hoc features, and of region labelling. We then turn to the modelling of a set of brain regions as a cooperative computation system--a distributed structure in which each system has its own 'goal structure' for selecting information to act on from its environment, and for transmitting the results to suitable receivers. We use this to describe a few findings of neurology. We then sample AI studies of computer understanding of natural language, ascribing particular significance to a speech understanding system configured as a cooperative computation system.

The literatures of AI and BT hardly overlap at all, and differ widely in choice of both problem and method. The aim of this article is to overcome these diversities by extracting contributions from extant AI and BT which can be melded into the creation of a top-down brain theory: the building of a coherent model of cooperative computation within which the computational roles of brain regions, and of neurons within those regions, can be analyzed.

1. INTRODUCTION

This article is a brain theorist's attempt to combine Artificial Intelligence (AI)--the art of programming computers to exhibit seemingly intelligent behavior--and Brain Theory (BT)--the study of the brain's function via mathematical modelling and computer simulation. Much of Brain Theory to date has tended to be bottom-up: working either from models of membranes up to neuronal function; or using simple models of neurons to predict or analyze the behavior of small or regular neural networks. Artificial Intelligence, on the other hand, has analyzed such processes as solving problems, planning, perceiving a visual scene, and understanding natural language. While many workers in the field of AI believe that introspection on the way they solve problems can help them program computers to solve those problems, a majority of such workers feel that the analysis of brain mechanisms involved in intelligent behavior is irrelevant to them. Rather, they take as their maxim "Airplanes do not flap their wings"--in other words, the detailed analysis of biological prototypes is no more relevant to the construction of robots than it is to the construction of airplanes.

While many problems in each of AI and BT will be of little interest or relevance to workers in the other discipline, this article will suggest that there is a conceptual overlap, and that workers in either field have much to gain from the effort to build a common framework. In particular, I shall be arguing the case for a top-down brain theory, which will use, modify, and contribute to AI techniques so as to better delimit the functions of brain regions in a form suitable for bottom-up BT analysis. Close to sensory receptors, we can correlate the activity of many neurons with stimuli; and close to the muscles, we can correlate neural activity with the animal's

responses. However, as we move away from these peripheral regions, such correlations with stimulus and response become almost meaningless. I shall argue that the goal of top-down brain theory can be rephrased as the building of a coherent model of cooperative computation within which the computational roles of brain regions, and of neurons within those regions, can be analyzed.

2. CONTROL OF MOVEMENT

In this section we analyze the control structures required for the movement of an animal or a robot. As we shall see, artificial intelligence studies have tended to focus on the high-level planning of various stages involved in an activity, while brain theory has had more to say about the regulation of posture and simple limb movements. We see here how a top-down approach from AI and a bottom-up approach from BT point the way toward a necessary synthesis.

2.1 Planning the Trajectory

Each side of the body is most strongly controlled by one half of the brain--with each hemisphere of the cerebral cortex talking to the opposite side of the body, while each half of the cerebellum has connections to its own side of the body. In a classic study of the role of the cerebellum in human movement, Holmes [1939] studied patients who had lost half of the cerebellum through a gunshot wound in World War I. Each patient was asked to stand in front of two vertical rows of lights and to move his hand swiftly back and forth. The normal subject, or the patient using the hand on the same side as the intact half of the cerebellum, exhibits a smooth trajectory of the kind shown in Figure 1a; while a patient using the hand whose cerebellum had been grossly damaged exhibits the erratic trajectory of

Figure 1b. One sees that the overall planning of the trajectory is unimpaired by the cerebellar damage--the patient still moves from one target to the next in the required sequence. What is lost, however, is the ability to predictively start decelerating the arm in time to halt the movement at the next target and initiate a smooth transition to the opposite direction. Studies of this kind have suggested that the cerebral cortex is involved in the high-level planning of an animal's movement, while the cerebellum is involved in the smoothing and modulating of the movement, including a predictive component. Actually, the situation is somewhat more subtle. In a basic activity like regular locomotion, the control algorithms seem to be located in the brainstem and spinal cord. But, here again we find that the movement is not as smooth or as well coordinated when the cerebellum of the animal is damaged.

Arbib, Franklin, and Nilsson [1968] were perhaps the first to relate this dichotomy of planning versus cerebellar modulation to the field of artificial intelligence. Their key observation was that many of Lashley's questions about "the problem of serial mechanisms in behavior" [1951] were answered as soon as one thought of the brain's computations not in terms of stimulus-response couplings, or simply chains of associations, but rather in terms of the types of planning activity which has been vigorously studied in AI. Before turning to a very brief review of some AI planning techniques, we first note that BT has had relatively little to contribute to the study of this high-level activity--though in Section 5.2, we shall discuss neurological studies which begin to shed light on necessary subsystems for the cooperative computation of high-level functions. Rather, much of the neurophysiological

analysis of movement has focused on spinal mechanisms--especially feedback mechanisms in posture and locomotion--and on single-cell correlates of stimulus or response. Clearly, one of the aims of top-down brain theory must be to better analyze the distribution of planning operations within cortical structures, and understand the signal flow that this planning must impose upon the cerebellum and other regions which modulate this planning.

We shall make clear the type of high-level planning involved in movement by considering two types of robots. A mobile robot is equipped with a television camera and wheels. The TV camera provides the input to a scene-analysis system, of the kind alluded to in Sections 3 and 4, which builds up an internal representation of the position and location of objects. The planning problem for this robot is to respond to a command such as "Push the large cube to the door" with a plan which will take it to the cube without bumping into obstacles, and then will allow it to push the cube to the door without bumping into any obstacles--and, then, to execute the plan. The other kind of robot is a hand-eye robot. It too has a TV camera which feeds a scene-analysis system which enables it to locate and classify objects. However, the robot has a mechanical hand as its effector, and a typical task for it would be to pick up two red blocks and stack one atop the other. Let us now turn to a brief discussion of the planning concepts involved here.

To simplify the planning problem for the mobile robot, we may schematize the world in which it finds itself by a graph. The nodes in this graph are targets for the movement of the robot--either places it must go to to push an object or pass through a door, or places it must go in skirting around an object so as not to bump into it. When given a task specified in

a high-level language, the robot must first express this in terms of conditions upon a path through the graph--such as to pass through certain nodes in a given order, while avoiding all nodes of another kind--and then it must find a path which satisfies these conditions and is also, ideally, as short as possible.

Clearly, the simplest form of this problem is to find the shortest path from one specified node to another without any condition upon the intervening nodes. Doran and Michie [1966] came up with a path-growing algorithm which used the idea of a heuristic distance, which indicates how far one has to go--but is only an indication, rather than a guaranteed measure. For example, when one is travelling by road, the distance by air is a heuristic distance for how far one still has to go. The Doran-Michie heuristic proceeds by exploring alternative paths, giving first preference to the exploration of paths whose next node has the least heuristic distance to the goal node. (It should be emphasized that this process of path-growing is made within the computer during the planning stage. At the completion of this path-growing process, a single path results, and it is this which directs the actual movement of the robot.) However, while this algorithm always yields a path from the start node to the goal node, it cannot be guaranteed that the path obtained is the shortest. This problem was overcome by Hart, Nilsson and Raphael [1968] who developed an algorithm which gave first preference to the exploration of paths whose next node had minimal sum of distance already traversed from the start node plus heuristic distance to the goal node. In later work, Fikes and Nilsson [1971] developed a system called STRIPS which--using theorem-proving techniques based on the representation of information about

the environment in terms of predicate calculus--could take account of the more detailed constraints on a path which must be followed by a mobile robot. Two sequel studies are of interest here. Fikes, Hart, and Nilsson [1972] studied generalized robot plans. Very briefly, the idea is to store information about the changes in the relationship between the robot and the environment brought about by a sequence of actions which are part of a successful plan. As time goes by, and other plans are made, various constants within both the sequence of operations and the prescription of its effects are replaced by variables in such a way that the robot has available a number of MACROPS--high-level operations which will let it meet many familiar planning problems. Sacerdoti [1974] has built on this study to define a problem solver called ABSTRIPS, which uses planning in a hierarchy of abstraction spaces. It usually proves far more efficient to plan the overall stages of a series of movements, and then refine the plan with greater and greater detail, only making modifications when necessary, than to grow the overall plan step by step, with each step defined at the lowest level.

The problems of the hand-eye robot stacking blocks allow us to emphasize a very important aspect of planning--namely, that it is often recursive. Consider how we might write a program for picking up a block. Before the block can be picked up, the lifting routine must call a routine for checking that the top is clear. But if the top is not clear, then this routine must in turn call the lifting routine--but now this lifting routine must be applied to the block, call it B_2 , which has been found to be atop the first block, call it B_1 . However, before B_2 can be lifted and placed elsewhere, we must check that its top is clear, and this may in turn involve the lifting of yet another block B_3 , and so on. Clearly, the problem of keeping track of the

nesting of calls to a few simple programs that this type of recursive procedure can entail can become very tedious. To handle the problem of recursion, and the problems of back-tracking when a tentative plan fails, Hewitt [1969] has devised a special programming language called PLANNER, which has been used by Winograd [1972], amongst others, in his well-known study of computer "Understanding of Natural Language".

2.2 Concepts from Control Theory

While artificial intelligence has recently upgraded the study of planning trajectories as a series of discrete steps, control theory has long looked at the optimal control of systems described by differential equations. Consider, for example, a space station. Its state is given at any time by its instantaneous position and momentum. Only a portion of this state--namely the position--is observable at any instant, and it requires cumulative observations of this output (position) over a period of time to build up a more and more accurate estimate of the velocity of the satellite, and thus of its momentum. There are inputs or control signals which can be applied to the satellite--for example, by adjusting the thrust of various rockets. A control system such as this satellite, then, can be expressed by a pair of equations

$$\dot{q}(t) = f(q(t), x(t)) \quad (1)$$

$$y(t) = \beta(q(t)) \quad (2)$$

The first expresses the rate of change $\dot{q}(t)$ of the state as a function of both the state $q(t)$ and the control vector $x(t)$ applied at any time t . The second equation expresses the current output $y(t)$ as a function of the current state $q(t)$. In the case of the satellite, β simply reads off the position vector

from the combination of position and momentum vectors which constitute the state.

The optimal control problem for any control system is then to transfer the system from some initial state q_0 to some desired final state q_1 in such a way as to minimize some criterion of the trajectory--such as the time, or the energy, or the fuel required to reach the final state. One of the basic observations of control theory--known as Bellman's principle of dynamic programming--is that the optimal control signal to apply to the control system at any time depends only on the instantaneous state of the system, and not on any further information about the prior trajectory. Thus the job of the controller can be divided into two parts:

- (1) The estimation problem: to build up from the history of observations of the output of the controlled system a better and better estimate of the current state of the system; and
- (2) The optimal control problem: to compute, on the basis of the estimate of the current state of the controlled system, the control signal which is optimal with respect to the given criterion in guiding the system towards the designated final state.

Both the state estimation and the optimal control computation must be based on precise values of the parameters which define the functions f and β in the system equations (1) and (2). However, in the real world the exact values of these parameters are seldom available. For example, even if a precise parametric description of a satellite is available on launching, this description will become less and less accurate as fuel is depleted, changing the distribution of mass. To adapt to such changes, the outer,

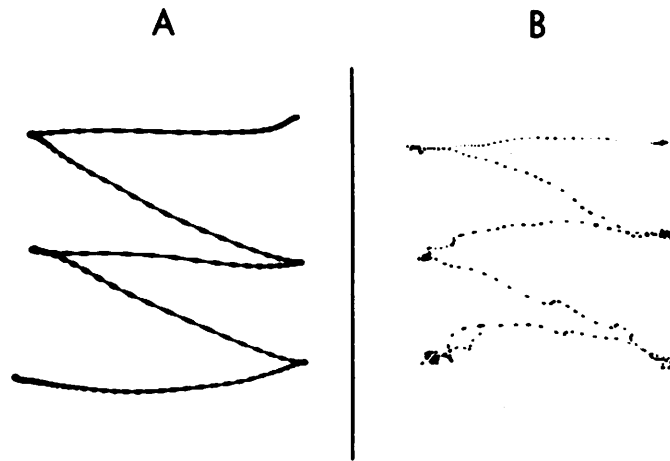


Figure 1:

Execution of back and forth motion by a patient using hand on the same side as (A) the intact half of the cerebellum; (B) the half of the cerebellum badly damaged by a gunshot wound. [Holmes, 1939]

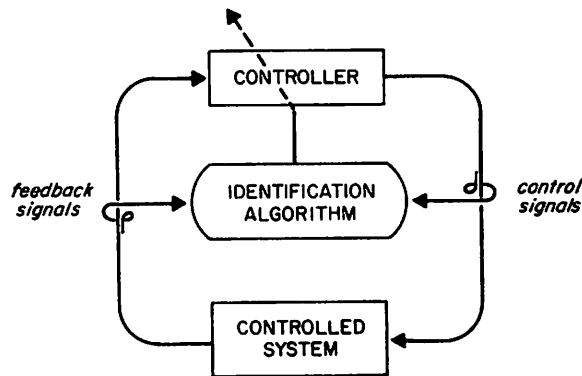


Figure 2:

To render a controller adaptive, an identification algorithm monitors control signals and feedback signals to provide the controller with updated estimates of the parameters which describe the controlled system.

feedback, loop of Figure 2 must be augmented by an identification algorithm. The job of this algorithm is to continually monitor the output of the controlled system and compare it with the output that would be expected on the basis of the current estimated state, the current estimated parameters, and the current control signals. On the basis of these data, the identification algorithm can identify more and more accurate estimates of the parameters which define the controlled system, and these updated parameters can then be supplied to the controller as the basis for its state estimation and control computations. Thus, with the addition of the identification algorithm we have a truly adaptive system.

Returning to the work of Holmes exemplified in Figure 1, we see that control theory studies the sort of function--providing the smoothest, shortest trajectory--which is so grossly impaired by damage to the cerebellum while the planning techniques of artificial intelligence seem to correspond to functions which are not impaired by cerebellum damage, namely the sequencing of overall subacts in such a way as to achieve the overall behavior. Brain theory has in fact addressed itself most to the neural implementation of control-type-mechanisms-within-a-step, rather than artificial-intelligence-mechanisms-for-putting-the-steps-together-in-a-high-level-plan. As already suggested, it is thus a major challenge to top-down brain theory to come up with a better specification of the contribution made by different brain regions in coordinating high-level planning and low-level modulation and coordination. This must be done in such a way that a neural network analysis can be conducted of the different regions, using realistic specifications of natural patterns of input and output activity.

2.3 Feedforward and the Cerebellum

Most bioengineers are familiar with the study of neural feedback mechanisms in the spinal cord--for a recent survey see Stein [1974]--but relatively few are familiar with the evidence for neural feedforward mechanisms. Whereas feedback uses the output of the controlled system to suggest to the controller which control signal should be applied to correct the system's trajectory, feedforward is used--as by the identification algorithm of Figure 2--to upgrade the controller's model of the controlled system, so that control may proceed with less error in the future.

To provide perspective, we first look at Figure 3, in which we see that the vestibular system--which monitors the acceleration, including gravitational forces, acting upon the organism--feeds the Deiters nucleus which, in cooperation with the cerebellum, can modulate the motor neurons in the spinal cord which control the contractions of the muscles of the limbs. It has long been known that massive stimulation of Deiters nucleus in the passive cat will lead to extension of the limbs, as a result of contraction of the extensor muscles. This makes biological sense, because massive activity of Deiters nucleus should signal massive activity in the vestibular system, and lacking the delicate patterning normally present in such input, the only interpretation for such a signal is, "Help, I'm falling over, but I don't know in which direction!"--to which extension of all limbs is the appropriate response.

However, it remained for Orlovsky [1972] to explore the effect of stimulating Deiters nucleus in the locomoting animal, rather than the passive animal in which the response to Deiters stimulation was a postural adjustment.

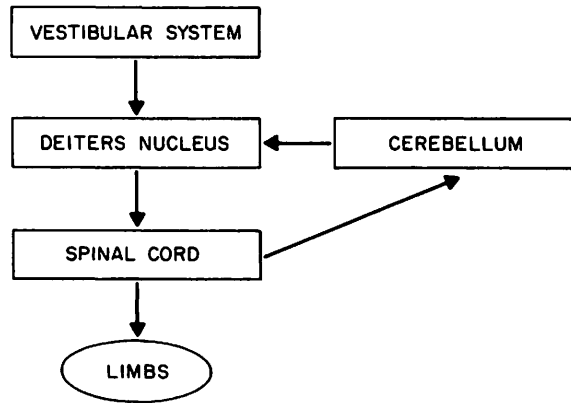


Figure 3:

The vestibular system, Deiters nucleus and cerebellum aid the spinal cord in postural adjustments of the limbs, and modulate spinal algorithms for locomotion.

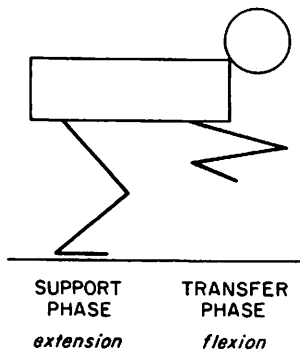


Figure 4:

The two major phases of stepping during locomotion.

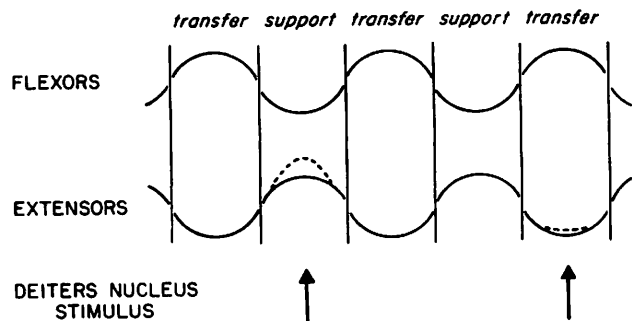


Figure 5:

The solid lines caricature the contraction of flexors and extensors during locomotion. The dashed lines schematize the findings of Orlovsky [1972] on the effect of Deiters nucleus stimulation on extensor contraction during locomotion.

Grillner [1975] reviews central mechanisms and reflex interaction in the control of locomotion in vertebrates. In describing the motion of a limb during locomotion, we can distinguish two gross phases (Figure 4): during the support phase, the limb is extended (by contraction of the extensor muscles) so that the foot is pressed firmly against the ground and pushed backwards so that, by reaction, the animal moves forward; whereas during the transfer phase the limb is flexed (by contraction of the flexor muscles) in such a way as to swing the leg forward, without touching the ground, ready to position it for the next support phase. The state of the muscle contraction can thus be caricatured as in the top two solid curves of Figure 5, with a high level of flexor contraction occurring during the transfer phase, and a high level of extensor contraction occurring during the support phase. Orlovsky [1972], then, stimulated Deiters nucleus--similar experiments were carried out on other nuclei of the brain stem--during locomotion. As could be expected from the stimulation experiments on the passive animal, no particular increase in flexor contraction was ever seen in response to Deiters nucleus stimulation. However, what was intriguing was the response of the extensors. During the support phase in which the extensors were actively contracting, Deiters nucleus yielded an increase in extension; whereas, during the transfer phase in which the extensors were not actively contracting, Deiters nucleus stimulation yielded virtually no increase in extension at all. This sort of study suggests very strongly that the role of Deiters nucleus is not--as the postural experiments might have suggested--to control the extension of muscles, but is rather to act as an identification algorithm, determining the appropriate parameter corresponding to degree of extension. In

the passive animal, the reigning algorithm in the spinal cord is the postural algorithm, and so a sudden increase in the extension parameter will indeed yield an extension response; whereas, during the reign of the locomotion algorithm, it is only during the support phase that we can expect the 'extension parameter' to be 'read out'.

The Moscow school of study of the control of movement founded by Bernstein--and of which Orlovsky is a member--has sought to analyze movement in terms of synergies--overall patterns of muscle contraction and relaxation which provide the 'vocabulary' of movement--the basic steps, as it were, which can be used as the building blocks in any planned behavior. An experiment like Orlovsky's suggests that there are regions of the brain whose job it is to provide feedforward adjustments of the various synergies involved in movement. This notion has been vigorously pursued by Boylls [1975] (see also Arbib, Boylls and Dev [1974]), who has modelled that part of the cerebellum--the anterior lobe--which is in communication with various nuclei in the brainstem which, like Deiters nucleus, can be implicated in the feedforward adjustments of synergistic parameters. He has shown that a wealth of anatomical and physiological data on the cerebellum may be encompassed in such a model of synergy adjustment. An account of his work, as well as of other studies of the cerebellum--including Ito's [1974] study of a possible feedforward role for the cerebellum in the vestibulo-ocular reflex--is given in Szentágothai and Arbib [1975]. For further information on the cerebellum, the reader may wish to consult Allen and Tsukahara [1974]; Eccles [1973]; and Llinas [1975].

It is beyond the scope of this tutorial article to go into details of the bottom-up analysis of neural circuitry in the cerebellum. Let us close this section, instead, by noting that while the studies of Orlovsky and

Boylls give us insight into the role of the cerebellum in providing feedforward adjustment for spinal control of locomotion, the data of Holmes shed light on the role of that portion of the cerebellum--the cerebellar hemispheres--which communicates directly with the cerebral cortex. In this case, we see that cerebellar damage impairs the feedforward which allows the animal or human to proceed confidently from a visual perception of the separation between two targets to a preprogrammed ballistic movement which can carry the animal's limb from one target to the next.

3. SCHEMAS AND ACTION-ORIENTED PERCEPTION

In this section, we provide a top-down analysis of various systems required in a perceptual system. Then, in Section 4, we shall provide bottom-up analyses--both from brain theory and from artificial intelligence--of mechanisms for a number of these, and related, systems.

3.1 The Slide-Box Metaphor

The Oxford English Dictionary defines "perceive" as meaning "to apprehend with the mind; to become aware or conscious of; to observe, understand." Leaving aside the problems of consciousness or awareness, we can say that the important element of perception is that aspects of the external world are "taken in" through the senses, and given an internal representation which must, however indirectly, be related to action if it is in some sense to contribute to the 'understanding' of the world that is being perceived. It seems plausible to suggest, then, that the act of perception must include the breaking down of the world into 'objects', where we define 'objects' as 'domains of interaction'. Thus depending upon the circumstances in which the organism finds itself, an 'object' could range anywhere from a marking on a stone, to a tree, to a whole row of trees whose detailed inter-relationships are not important to the animal's current behavior. Thus, if we think of the world as an array of domains of interaction, then it becomes appropriate to think of the internal representation of the world as being an array of representations of 'objects'. In trying to make this idea more vivid, Arbib [1970; 1972] introduced the slide-box metaphor. In the making of moving-picture cartoons, it is far too much work to draw every frame anew.

Rather, the background remains unchanged through many frames and can be drawn once; the middle ground may not change anything save its position relative to the background, and so can be drawn on a transparency and simply repositioned in front of the background from one frame to the next; while, even in the foreground, only some details need to be redrawn to express current activity. In the slide-box metaphor, then, we think of a 'slide' as being the internal representation of an 'object', so that the current 'scene' is represented by a 'collage' of slides. What is important is that this collage is dynamic, and requires relatively economical updating over time as the stimuli to the organism change. In response to both internal goals and needs, as well as to sensory stimulation, slides can be tuned, relocated, and replaced. The utility of a 'slide', in this metaphorical sense, is that it is not simply the piece of celluloid of the technology which gave rise to the metaphor, but is rather a structure which gives access to programs appropriate for interaction with the 'object' that the slide represents. In Section 3.2, we shall move from this highly metaphorical presentation to a revised slide model of internal representation. We shall see in Section 3.3 that this model involves heterarchical cooperative computation by autonomously active schemas (!). But first, we should look at some related concepts.

From psychology, we have the notion of scheme due to Bartlett [1932], and the more elaborate theory of sensorimotor schemas due to Piaget (see Furth [1969] for an exposition) in which sensory information is used to cue appropriate motor activity. Another psychological approach, explicitly based on the early AI work of Newell, Shaw, and Simon [1960] (for an overview of which see Newell and Simon [1972]), is the theory of 'plans' of Miller, Galanter, and Pribram [1960]. They talk explicitly of infor-

mation being used to get access to appropriate programs--or plans--which can then be followed to completion.

A number of workers in artificial intelligence have also developed ideas of internal representation--with much the same ancestry as the ideas of Miller, Galanter and Pribram. With increasing interest in getting computers to understand natural language (see Schank and Colby [1973] for some recent contributions), workers in AI have not only looked at the representation of such linguistic knowledge as syntax and semantics, but have also concerned themselves with the representation of the knowledge of social interactions required if a computer is to understand a story (Charniak [1975])--here is an example cited by Minsky [1975]:

Jane was invited to Jack's birthday party.

She wondered if he would like a kite.

She went to her room and shook her piggy bank.

It made no sound.

Minsky observes that no sense can be made of this simple story unless one has available a representation--he calls it a 'frame'--of knowledge about birthday parties, etc. One has to know that birthday parties require presents, that presents require money, that children get money from a piggy bank, and that one can test whether there is money in a piggy bank by shaking it, and that if there is no sound there is no money. Similar work has been done by Schank and Abelson [1975] in their study of plans and scripts, and by Schmidt [1975] in his work on belief systems, which can be seen in some ways as a successor to the study of "Speech Acts" by Searle [1965]. In this context of the representation of social knowledge, it is intriguing to notice the recent publication of a book entitled Frame Analysis: An Essay on the Organization of Experience. It is not an account of Minsky's work,

but is rather the attempt of a social psychologist, Goffman [1974], to represent the different types of context which determine the way in which people behave. For example, he notes that when you walk into the doctor's office and the doctor asks, "How are you?" you reply, "Fine, thanks"; and yet, when you are both seated, and he asks you again, "How are you?" you reply with a detailed list of troublesome symptoms. What has changed? Goffman suggests that all that has changed is that you have moved from the 'greeting' frame to the 'doctor-patient' frame. For a discussion of the relation between frames and the approach espoused here, see Arbib [1975b].

3.2 The Definition of a Schema

In trying to put the suggestions for a theory of internal representations expressed above in a more formal framework, Arbib [1975a] has suggested that the metaphorical concept of a 'slide' be replaced by the more formal concept of a schema. This is a system with three components:

(i) Input-matching routines: A schema corresponds to an 'object' (though this may be at a very abstract level, such as 'winter', or 'a differential equation') and so requires routines whose job it is to search sensory stimuli, as well as messages from other schemas, for cues as to the correctness of the hypothesis that the 'object' which the schema represents is indeed present in the system's environment. These input-matching routines cooperate with segmentation routines which are continually trying to segment the world into analyzable regions. It must be stressed that the input which is being matched may be multi-modal--one may perceive a cat as much from its meow or the feel of its fur as from seeing it. Again, the level of detail at which the input is matched may vary greatly with the goal structure of

the animal. If one is looking for a suitable stick, the detailed branching structure of a tree must be perceived; whereas if one is simply taking a walk, it may be enough simply to perceive 'a-row-of-trees' and avoid bumping into it. Finally, it must be stressed that the input-matching routines, in their fullest generality, will not simply match static aspects of the environment, but will match dynamic aspects--as, when crossing the road, one is more interested in perceiving those dynamic aspects of a car required to avoid being hit than one is in the make of the car.

(ii) Action routines: As we have stressed from our action-oriented viewpoint, a schema must include action routines, appropriate for guiding the activity of the organism in interacting with the 'object' which the schema represents. As input-matching routines adjust the parameters of the representation more accurately, the action routines should be adjusted so that the action they would release becomes more and more appropriate for the current environment and goal structures.

(iii) Competition and cooperation routines: As we shall spell out in more detail in Section 3.3, different schemas will compete to 'cover' the 'object' in a given part of the animal's world; while other schemas will cooperate to between them provide a coherent representation of a number of regions in the world. However, even when, through this process of competition and cooperation, a 'collage' of active schemas has provided an acceptable representation of the environment, one problem still remains: the animal cannot at any one time carry out more than a small fraction of the possible patterns of interaction consistent with the action routines of the activated schemas. We may say that the active schemas present, in McCulloch's phrase, a redundancy of potential command--each could, in principle, command the

activity of the organism, but there are too many of them to command simultaneously. Thus, further competition and cooperation routines are required to turn the range of possibilities into a coherent plan of action for the organism.

In an extended theory of schemas, one must not only spell out, for example, the detailed working of the competition and cooperation routines, but must specify how increased effectiveness of input-matching serves to tune the action parameters; specify the way in which the matching of dynamic properties of 'objects' enables the organism to act in a predictive fashion; and specify the way in which the organism can 'learn from experience'. This updating of memory structures must involve the combining of old schemas to form new schemas; the tuning and editing of schemas to better fit them to a changing world; and provision of increasingly rich relational information--embodied in part in competition and cooperation routines--to coordinate schemas to better encode relations between the 'objects' that they represent. In the following subsection, however, we shall turn our backs on this richness, and outline two problems, those of segmentation and of region labelling.

3.3 Segmentation and Region Labelling

We have stressed the structure of the internal representation as a 'collage' of representations of 'objects'. The problem of segmentation is precisely this problem of breaking the input into meaningful pieces--whether it be perceptibly distinct objects in a seen scene; or perceptibly

distinct words in a heard sentence (see Section 5.3). In this subsection we will focus upon the visual problem.

As one looks around a room, one can 'break' the visual world into regions on a variety of 'low-level cues'--such as depth, color, and texture. A region may stand out from other regions literally (depth cues), or because it is of a relatively homogeneous color or texture which differentiates it from surrounding regions. A region that has been 'segmented out' in this way can then provide excellent input for a 'higher-level' routine which attempts to determine the 'object' of which this region is a part. We shall look at detailed models of segmentation on depth cues in Section 4.3 and of segmentation on color or texture cues in Section 4.4. What should be emphasized in this introductory statement is that the processes of segmentation on low-level features and of 'input-matching' to determine the 'object' of which the segment is a part, do not proceed in the fixed order: segmentation first, 'object assignment' second. Rather, it will often be the case that when one region is segmented, it will provide a cue which greatly increases the activity of a particular schema. The high activation level of this schema will suggest the presence of another region having certain properties, and this hypothesis can then speed up the low-level segmentation processes in that region. And so it goes, in a highly cooperative process.

Rather than trying to provide at this stage a detailed model of how a schema can 'cover' a number of different regions in a visual scene, and in so doing tune parameters which prepare the animal for interaction

with the 'object' represented, we look at a stripped-down problem, namely that of region labelling. Consider Figure 6. One can think of at least two hypotheses which are consistent with the decomposition of this scene into regions, the top of which is green, and the bottom of which is brown. One is that we are viewing a pistachio ice cream (!); the other is that we are viewing a tree. In terms of competition and cooperation routines between schemas, we may then imagine that, amongst the array of schemas shown in Figure 6, the top region serves to equally activate the foliage and ice cream schemas, while the brown region serves to equally activate the trunk and cone schemas. There is excitatory interaction between foliage and trunk schemas, and between the ice cream and cone schemas while each of these pairs inhibits the other pair. Without further information, one cannot confidently predict convergence of the array of schemas upon a state in which one pair rather than the other is active. However, if extra information is provided, as in Figure 7, by context which preferentially activates a hand schema rather than a grass schema for a region adjacent to and below the brown region, then one can expect that the positive interaction between the hand and the cone schemas coupled with the negative interaction between the hand and the trunk schemas will between them serve to move the system into a stable state in which the ice cream cone hypothesis is strongly favored over the tree hypothesis. In Section 4.4, we shall put this informal description on a more rigorous basis, and do that within a general framework of competition and cooperation networks which also embraces powerful mechanisms for segmentation.

As a brain theorist, I view the schema as being autonomously active. Each one is continually monitoring input from periphery and from

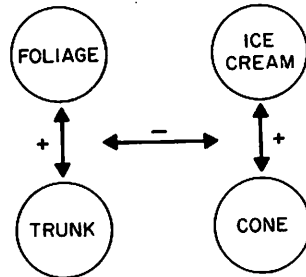
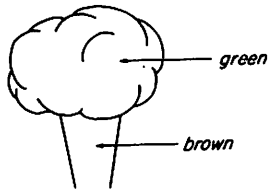


Figure 6:

The 'tree' and 'ice'cream cone' hypotheses compete for the input picture.

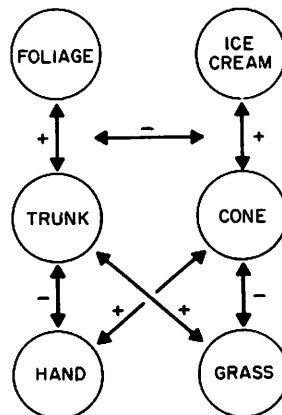
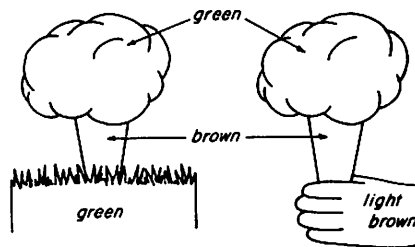


Figure 7:

Context can tip the balance between the hypotheses of Figure 6.

other schemas, trying to increase its own activity and influence the activity of the other schemas. Out of this cooperative and competitive interaction emerges a stable pattern in equilibrium with the environmental input and which constitutes the representation of the input. Of course, when one tries to simulate such a system on the computer, or if one is doing straight scene analysis within an AI context, one will try to find an implementation of this process in terms of a serial program, and thus emphasis will be on exploring one hypothesis at a time, and back-tracking when that hypothesis fails. Even then, though, it may be a mistake to think of this back-tracking as being the essential logic--rather it is the cooperative computation by autonomously active schemas that is the real logic of perceptual activity, even though any limited system--be it a computer or a brain--must in some way cut down the totality of interactions.

It should be stressed that there is no strict hierarchy of (1) activating low-level features, (2) activating segmentation, (3) labelling regions, (4) lumping labelled regions into 'objects', etc. Rather, as Figure 8 suggests, there is a heterarchy of interaction. Strictly speaking, heterarchy means "rule by an alien leader", but, following the usage of McCulloch [1949], it is now used in the AI community to indicate a situation in which A can command B at one stage, while B may command A at another stage. For example, in looking at a picture, the recognition that a white region is snow may be the sign for the recognition that this is a winter scene--in this case, the 'snow schema' feeds the 'winter schema'. However, if one is looking out the window to see what the weather is like, one knows full well what the season is,

and the knowledge that it is indeed winter can--if one lives in that part of the world--trigger the recognition that the white region out there is snow rather than moonlit water or a stretch of burnished sand.

In closing this section, let us just stress once again the generative nature of the internal representations we have posited: the representation of the world at any time is a 'collage' of active schemas. Thus, the animal can internally represent situations which it has never encountered before because the elements of the situation are familiar, even though the combination of those elements is completely novel to the organism.

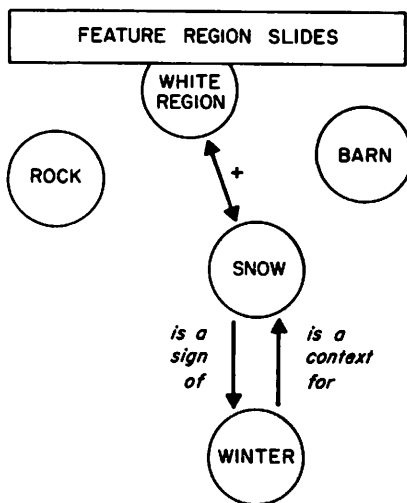


Figure 8:

A heterarchy of interaction between schemas.

4. COMPETITION AND COOPERATION

In developing our top-down theory of schemas as the internal representations of 'objects' for a system with action-oriented perception, we focused upon two problems: the segmentation problem of dividing the input into meaningful regions for further analysis; and the region-labelling problem, of deciding between alternative labellings for the regions so obtained. We also emphasized that this 'straight through' version was too simplistic. However, our task in this section is to provide mechanisms for these separate processes, and to do so in a framework which provides many other bottom-up models of brain mechanisms involving competition and cooperation.

4.1 From Executive to Cooperative

In suggesting a scheme for the recognition of letters, Selfridge [1959] posited the "Pandemonium" structure shown in Figure 9a in which a collection of demons each looked at the input data, and on that basis determined a "confidence level" in the hypothesis that the input was the demon's own letter. In this sense, the demons are much like the schemas of Section 3. However, instead of competition and cooperation routines being used to determine the 'top candidate' to 'cover' the input, Selfridge posited an "executive demon" whose job was to listen to the "yelling" of the other demons, and determine which demon yelled the loudest--this demon's decision was then that adopted by the "executive demon".

One of the first models of decision making in neural circuitry to

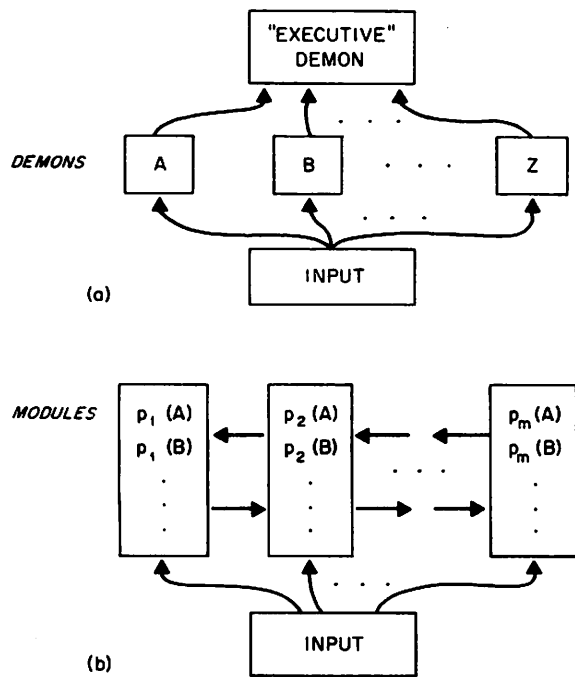


Figure 9:

Selfridge's Pandemonium model (a) requires an executive to resolve conflict; while in the S-RETIC model (b) of Kilmer and McCulloch, interaction between modules yields consensus without executive control.

explicitly opt for cooperative computation, rather than executive control, was the S-RETIC model of Kilmer, McCulloch and Blum [1969]. In modelling the reticular formation of the brainstem, they used observations that the reticular formation helped switch the organism's gross state from sleep to wakefulness, and vice versa, to suggest that the function of the system was to determine the overall mode of behavior of the organism--such as sleeping, fighting, fleeing, or feeding. From the anatomical work of the Scheibels [1968], they felt that a reasonable structural simplification of the system was as a stack of "poker chips". In their model (Figure 9b), each module in the stack was to receive a sample of the overall system input, and on the basis of that sampling to assign weights to the different modes--so that at a given time, $p_i(\lambda)$ was to denote the weight given by module i , based on the system's input, to the hypothesis that the appropriate mode of the organism was λ . However, the modules were then to be coupled in such a way that each module would readjust its weights on the basis of activity in other modules to which it was connected. Kilmer and McCulloch were able to suggest a connection scheme which would lead to eventual consensus, with a majority of the modules assigning the greatest weight to a single mode. This is all done without any executive control.

A possible analogy is that of a panel of physicians, each a specialist so that he pays most attention to symptoms relevant to his speciality, and yet with enough general medicine to be swayed by the comments of the other specialists. By talking back and forth amongst themselves, the panel is able to reach consensus as to the diagnosis of the patient, without the necessity of having any 'chief physician' tell them what they must do. As we shall see below, S-RETIC may be regarded as the prototype

for a whole class of general models of competition and cooperation which include mechanisms for segmentation and region labelling. In the next two sections, we shall present two other specific models of competition and cooperation, and shall then examine the general framework in Section 4.4.

4.2 From Frogs to Human Eye Movements

One of the best-known results of neurophysiology is the observation by Lettvin, Maturana, McCulloch and Pitts [1959] that the ganglion cells of the retina do not simply send back a light intensity map to the brain, but rather transmit preprocessed information. In particular, one class of ganglion cells has been called "bug detectors" because these cells fire most vigorously when a stimulus wiggling like a fly is applied in a particular region of the visual field. Ingle's [1968] study of the snapping behavior of the frog provides data for an action-oriented analysis of how the animal makes use of these neural signals. Of most interest here is that in a certain region around the head, the presence of a fly-like stimulus will elicit a snap from the frog--that is, it will turn so that its midline is pointed at the 'fly', and the frog will zap it with its tongue. When confronted with two 'flies' either of which were vigorous enough that alone they would have elicited a snapping response (Figure 10), the frog could exhibit one of three reactions: it could snap at one of them; it could not snap at all; or it might snap at the 'average fly'. From the viewpoint that we have built up in Section 3, it is as if each fly-like stimulus is sufficient



Figure 10:

A frog confronted with two "flies".

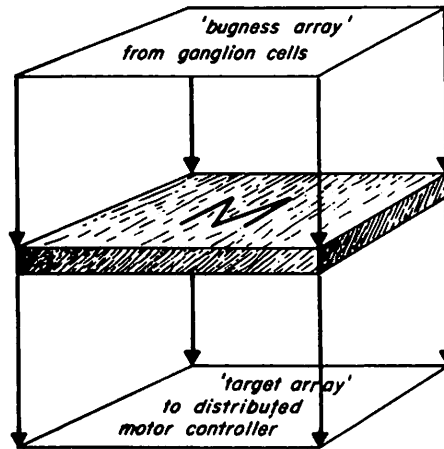


Figure 11:

The task of the competition network, M, is to normally let through only one strong signal from the layer of bug detector signals to the control surface of the distributed motor controller.

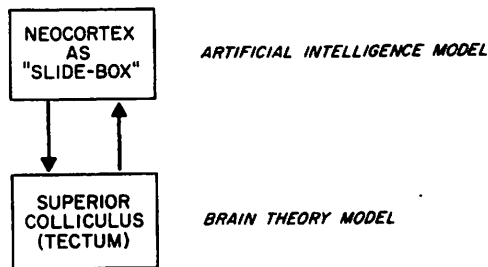


Figure 12:

A hybrid AI/BT model for the study of eye movements and visual perception.

to activate a simple 'fly schema', and that each fly-schema has associated with it an action routine for snapping in the appropriate direction. However, there is competition between the action routines of these schemas.

It was Didday [1970] who gave a distributed network solution to this problem of competitive interaction. He modelled the tectum (Ingle and Sprague [1975])--the visual region of the midbrain which receives information from the ganglia, including the outputs of the bug-detectors. From the work of Braitenberg and Onesto [1960] and Pitts and McCulloch [1947], I had posited that the tectum must feed a distributed motor controller, with activation of several points on the surface of this controller eliciting a snapping movement of the frog at the center of gravity of the corresponding locus in space (Figure 11). The task of the competition network, then, was to normally let only one strong signal through from the layer of bug-detector signals to the control surface of the distributed motor controller. However, Ingle's experiments suggested that this process could fail in two realistic ways: in one case the competition between two strong inputs would be such that each would hold the other down and neither would become strong enough to take control of the system; while in another case two signals which were properly balanced could get through simultaneously, yielding a movement to their 'mean'. The essential mechanism here was to provide each region of the tectum, corresponding to a region of the visual field, with a 'sameness' cell, whose job was to come up with a weighted sum of activity elsewhere and use this to diminish activity within its own region. In this way, all but the strongest loci of activity would be suppressed, and these activities would then be able to 'come up out of the noise' to take control--

save in the case where two regions were sufficiently strong to each hold the other down below the threshold for motor control.

At this stage, a few remarks on methodology may be in order. It should not be thought that the snapping of frogs at flies is considered to be an end in itself--even though it can become quite fascinating once you 'get into it'! Rather, the brain scientist is motivated by the comparative approach--he hopes to learn about the brains of humans by comparing them with the brains of other animals. In particular, the tectum of the frog, which we have just modelled, is homologous to the superior colliculus of the human. In other words, these two midbrain visual regions have much in common both anatomically and functionally. In a brain much closer to the human's, Wurtz and Goldberg [1971] have indeed found that close to the top of the monkey's superior colliculus, the response of the cells seems best correlated with the stimulation to the animal, whereas in the deepest layers the correlation is far better with the movement of the animal. Thus, it is plausible to believe that an adequate model, such as Didday's, of the transformations which take place as we go down through the frog tectum will yield insight into the transformations taking place in the superior colliculus of monkeys--and of humans. Where the visual input in the frog triggers the snapping at an object, the visual input to the superior colliculus can trigger an eye movement--as when we look at the source of a sudden movement or a flash of light. Didday and Arbib [1975] have built on this sort of observation to come up with a model of the role of eye movements in visual perception. They use the language of Section 3 to model the neocortex as a 'slide-box'--in other words

a system in which an array of schemas can build up representations of the world; while they model the superior colliculus in essentially the way in which Didday modelled the frog tectum. Thus, the model of the neocortex is at a top-down, almost AI level; while the model of superior colliculus is at neural network level, at the level of bottom-up BT (Figure 12). Essentially, the idea of the model is that the superior colliculus not only receives the direct input concerning the sudden flashes of light, etc., which can distract attention; but also receives 'calls' from neocortex--as when a schema needs more detailed input from the fovea to confirm a hypothesis about a particular region; or when there seems to be a discrepancy between what the schema has predicted and what is actually happening in a region of periphery, which thus demands more scrupulous attention. The idea of such a hybrid AI/BT model is that one can hopefully use the posited interaction to suggest experiments upon the signal flow between two regions, even when one of them does not yet have a neural net model; and that, to the extent that these experiments succeed, one can then more tightly constrain that region to bring it closer to a neural net analysis.

Many people have discussed the question of the extent to which the brain is a serial or a parallel computer. The type of model we have just looked at suggests that this is a spurious distinction. In looking at eye movements, we see a series of events taking place, but we realize that within each episode, highly parallel computation is required. Given the amount of parallelism we have posited in the visual system, it might well be asked, then, why we require a series of visual fixations. To the

answer that we only have a relatively small fovea with high visual acuity, it can quickly be responded that it would presumably not be too great an embryological trick to reiterate the foveal pattern over a much wider area of the retina. However, the real reason may be the following. Imagine that there were a superfrog which shares the frog's ability of having the necessary 'schema machinery' reiterated for every region of the retina, but that shares the human ability of having a wide variety of schemas available to guide all manner of sophisticated interactions with the environment. The trouble with this arises if the organism sees something fearsome that it has never experienced before. Not knowing that it is fearsome, it gets into serious trouble with it before the encounter is over. If superfrog has the recognition machinery replicated for every region of the retina, it must then ensure that that experience is communicated to every such region and appropriately stored. As the world of the organism becomes more and more complicated, the combinatorial explosion required by such communication would become a far higher price to pay than the price paid by requiring eye movements for successive foveation.

4.3 Segmentation on Prewired Features

As we remarked in Section 3.3, a visual scene can be segmented on many different features, such as depth, color, and texture. In this section we give a model of segmentation on depth cues, and show that it can be placed in a common perspective with the Kilmer and McCulloch model

of S-RETIC that we studied in Section 4.1 and the Didday model of the frog tectum that we studied in Section 4.2. Then, in the next section we shall proceed further with formalizing our study of schemas by looking at a model of segmentation on color and texture cues, and a model of region labelling. Intriguingly, the latter model will be seen as a straightforward generalization of S-RETIC.

Turning, then, to the problem of segmentation on depth cues, we may note that depth in space is signalled by disparity of retinal projection. Whereas (Figure 13) the focal point corresponds to projections with the same coordinates on each retina, we can see that as we move back and forth along the ray AB, different points of the left retina are paired with the projection F on the right retina. It is this difference of the retinal coordinates of the two eyes that we call the disparity. Barlow, Blakemore and Pettigrew [1967] and Pettigrew, Nikara and Bishop [1968] found that in fact cells in visual cortex are tuned for disparity. More precisely, the line detectors of Hubel and Wiesel [1962] do not only respond best for a particular orientation and spatial direction of a stimulus, but respond best when the disparity of the projection of that preferred orientation upon the retinae is within some particular range.

Turning from this physiological evidence of the existence of 'depth detectors' in visual cortex to some psychological data on depth perception, we may note the dramatic work of Julesz [1971]. He prepared input to the two eyes in which each eye received a totally random pattern, but in which there were correlations between the inputs to the two eyes. Specifically, different regions in the two inputs were identical save for a shift in position, yielding--then--a different disparity on the

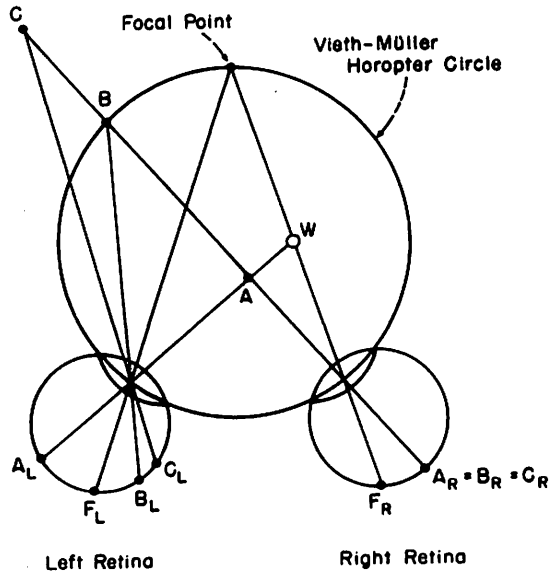


Figure 13:

Disparity of retinal projection is a sign of the distance along a ray to an external point.

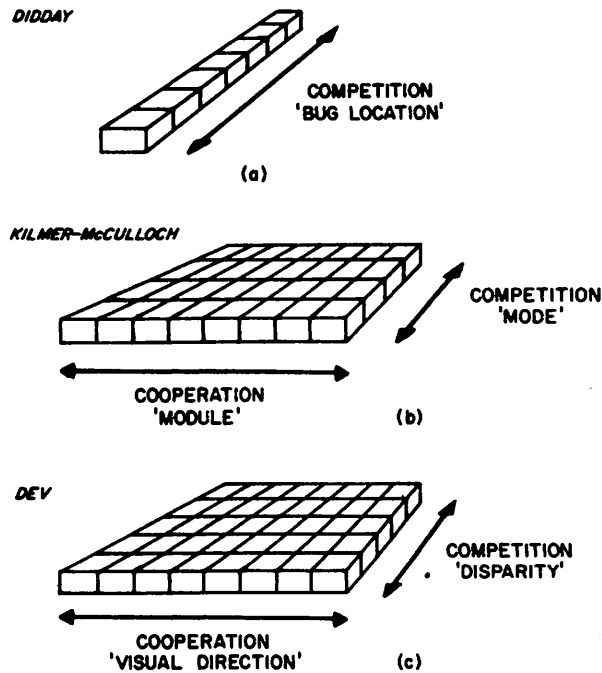


Figure 14:

The Didday model of frog tectum (a), the Kilmer-McCulloch S-RETIC model (b) and the Dev stereopsis model (c) set forth in a common framework of competition and cooperation dimensions [Montalvo, 1975].

retinae. Although such a pattern for a naive subject can initially appear to be nothing but visual noise, eventually the disparity matching takes place and he perceives surfaces at different depths. Presumably, what causes the initial perception of noise is that in addition to the correct correlation of points in the two retinae there are many spurious correlations, and computation is required to reduce these spurious correlations.

The solution to this proposed by Dev [1975] (see Arbib, Boylls and Dev [1974] and also Nelson [1975]) was to imagine the cells of a given disparity as forming a population arrayed in a spatial map corresponding to the map of visual direction. Connections between cells were then to be arranged so that nearby cells of a given disparity would be mutually excitatory; whereas cells which were nearby in visual direction but different in disparity would have inhibitory interaction. In this way, the activity of the array would become organized into a pattern in which in each region of visual direction cells of only one disparity type were highly active. As a result, the visual input would eventually be organized into a number of distinct surfaces. This scheme can be represented as in Figure 14C, in which we use a one-dimensional representation of spatial directions, and then have a second dimension for disparity. We have competition along the disparity axis, and cooperation along the space axis. We can immediately note, with Montalvo [1975], that the same diagram, Figure 14B, describes S-RETIC as well: save that the cooperation dimension is now the array of modules, and the competition is between modes rather than disparities. Finally, we can see, in Figure 14A, that the Didday model can be regarded as the limiting case of this general scheme of competition and cooperation, where now there is only a competition dimension, namely that of bug location.

Placing these particular models in this common setting stresses the need for a general mathematical framework (which is currently being worked on). What are the conditions of homogeneity, and the limitations on the length of the cooperation dimension, that guarantee that S-RETIC will converge to a state in which there is consensus in the sense of a single mode capturing the highest activity for a majority of the modules; whereas in the Dev model one has nonhomogeneous input, and the goal is to divide the cooperation dimension into regions, in a manner dependent upon the input, with each region having activity at a single locus on the competition dimension? As we shall see in the next section, the region labelling problem again fits into this general framework, save that the cooperation dimension has a topology determined by the spatial relationship between the regions to be labelled, and the goal now is a consistent labelling, rather than consensus within any one neighborhood of regions.

4.4 Back to Schemas

While it is plausible that there are prewired feature detectors for a variety of different disparities, it is certainly not plausible that there exists a spatially-tuned array of feature detectors for every color cue or every texture cue that a human could use in segmenting visual input. A very significant contribution, then, is a scheme for segmentation on ad hoc features. One such scheme has been developed by Hanson and Riseman in their work on a preprocessing system for a robot which is to recognize outdoor scenes. They work with color features, so that the raw image is

composed of three intensity maps in red, green and blue. A microtexture is then a color pattern detectable in a local window and which is repeated with minor variations over many such windows; while a macrotexture is an iterated pattern of spatially related microtextures, as in the interplay of green foliage and dark shadow in a tree during the summer; or the interplay of brown branches and blue sky in a tree during winter. Their scheme (Hanson, Riseman, and Nagin [1975]) for segmentation on ad hoc macrotextures is suggested in Figure 15. The raw image is preprocessed along a number of feature dimensions, and a histogram is constructed in feature space--so that there will be a clustering of points in feature space for those feature vectors which describe many windows in the original raw image. This histogram, then, is used as the basis for a clustering algorithm, and each of these clusters is taken to indicate a microtexture in the original image. Each window in the image can then be labelled with the name of the cluster to which its features belong. An adjacency matrix is then used, along with other techniques, to detect common patterns and adjacency between macrofeatures, and these adjacency patterns then provide the macrofeatures of the image. Finally, with this vocabulary of ad hoc macrofeatures, segmentation can proceed to give an image which is segmented on features which may never have been seen before by the system, even though those features are built up from the wired-in preprocessing of the system.

Having seen how to segment on both prewired and ad hoc features, we now turn to a scheme due to Rosenfeld, Hummel and Zucker [1975] for region labelling. As we have remarked before, the structure is very much like that of S-RETIC, but here the goal is a consistent labelling of regions,

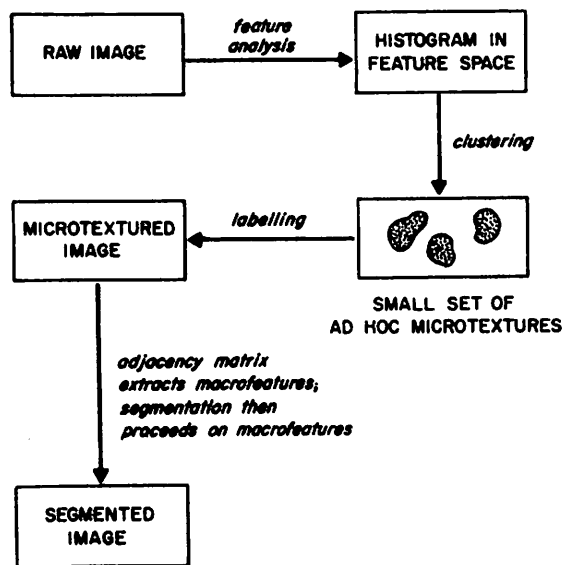


Figure 15:

The Hanson-Riseman scheme for extraction of ad hoc features for segmentation.

rather than a consensus. The reader may find it worth referring back to the ice cream vs. tree example of Section 3.3.

Formally, then, we have a set $A = \{a_1, \dots, a_m\}$ of regions, and a set $\Lambda = \{\lambda_1, \dots, \lambda_n\}$ of labels. The initial preprocessing will assign weight $p_i(\lambda)$ to the hypothesis that λ is the proper label for a_i . The problem now is to adjust those weights. This is done by using expressions $r_{ij}(\lambda, \lambda')$ in $[-1, 1]$ for the compatibility of λ on region a_i with label λ' on a_j . For example, if it is known that region a_i is above region a_j , then $\lambda = \text{foliage}$ will be highly compatible with $\lambda' = \text{trunk}$, and so $r_{ij}(\lambda, \lambda')$ for these values will be positive and close to 1; whereas if one reverses λ and λ' one would expect the compatibility to be far smaller. The idea is to adjust weights as follows: if another region has assigned high weight to a label which is compatible with a_i bearing label λ , then $p_i(\lambda)$ should be increased; whereas if high weight has been assigned to a region that is incompatible, then $p_i(\lambda)$ should be decreased. This suggests that the appropriate change to make in $p_i(\lambda)$ is

$$\Delta p_i(\lambda) = \sum_j d_{ij} \sum_{\lambda'} r_{ij}(\lambda, \lambda') p_j(\lambda').$$

for suitable coefficients d_{ij} (with $d_{ij} \geq 0$, $\sum_j d_{ij} = 1$). In other words, our first pass would be to replace the vector p_i with the vector $p_i + \Delta p_i$. The trouble with this, of course, is that the result is not a probability vector, and so we must pass this sum through a nonlinear operator R that produces the desired probability vector. Thus the next estimate of the weights is given by the expression

$$\hat{p}_i = R[p_i + \Delta p_i].$$

In fact, except for the choice of r_{ij} 's, this expression is almost exactly that given in the formal development of S-RETIC. Thus we see that we have in this section a powerful general procedure which crops us again and again both in such top-down problems common throughout artificial intelligence and brain theory as region labelling; and such models of bottom-up brain theory as the modelling of the frog tectum. We close by noting one other low-level example: the regions might in fact be rather small windows of the visual field, and the labels might be orientations of lines within those regions. The compatibility factors would then be non-zero only for adjacent regions, and would be such as to strongly favor line orientations in the two regions which would line up to form part of a continuous curve. The process given above, operating on these data, would then serve to grow continuous curves on the basis of the raw data.

5. COOPERATIVE COMPUTATION

In previous sections, we have explored the idea of cooperation and competition in networks for segmentation, mode selection and region labelling systems. In this section, we study a higher-level notion of cooperative computation, suggesting that the brain is a distributed structure of cooperating systems. In the first section, we shall spell out in more detail this notion of cooperative computation. Then, in Section 5.2, we shall suggest that the proper approach to systematizing the findings of neurology--which is essentially a top-down approach--is to model a set of brain regions by a cooperative computation system, and then correlate lesion induced defects of the brain with subsystem deletion. Finally, we shall look at some approaches to linguistics now current in AI, and observe that although only one of them is structured in the style of cooperative computation, there is good reason to believe that this style will in fact become the accepted one. It should, however, be noted that many AI workers do not yet hold this view.

5.1 Contracts and Hierarchies

We have suggested that we model a set of brain regions as a cooperative computation system--a distributed structure in which each system has its own 'goal structure' for selecting information to act on from its environment, and for transmitting the results to suitable receivers (or simply broadcasting them). To this we must add that in many realistic systems, the overall behavior must be produced within a limited time and that there is often no right answer--rather the best answer available within

the time limitations. Thus, rather than debugging a system in the sense of rewriting it to remove departures from an overall prespecification of its behavior, we must restructure the system on the basis of comparisons of "result y produced in context c with input x " and "result y' produced in context c' with input x' " to ensure that over time the overall behavior of the system increases cumulatively.

Since the main technique of neurology is to learn about the structure of a system by observing lesion-induced defects, we shall want to model the effects of subsystem deletion in a distributed structure of cooperating systems. In this context, it may be worth recalling the view espoused by the Russian neurologist Luria of localization as "a network of complex dynamic structures, of mosaics united in a common task with successive and simultaneous activity in a number of areas required to achieve any particular function." Such a system must be handled carefully. Gregory [1961], for example, has observed that if one removes a resistor from a radio, and the radio then emits a howling sound, one is not justified in calling the resistor a Howl Inhibition Center. This, of course, is a lesson well-known to all neurologists, and so they consider the effects of many different lesions upon replicas of a given system in seeking consistent hypotheses about the subsystems so as to build successive approximations to the structure of the overall system. Lashley's [1929] work has often been adduced against the idea that the brain is a mosaic of precise subsystems. He removed portions of cortex, and found that the decrease in performance in rats running a maze did not seem to depend on where in non-sensory cortex the removal had been made [the law of equipotentiality], but that the defect did seem to increase with the amount of tissue removed [the law of mass action]. However, it has since become well established that

one cannot infer from this that the brain is a "diffuse system", since the task is very simple for the rat and different lesions impair the performance of this simple task in different ways.

Let us look at this matter of subroutine deletion in more detail. A normal serial computation will crash if any instruction, let alone subroutine, is removed. In the type of parallel computation embodied in a patch of neural network, one may use redundancy to correct any localized errors--as in the work of Winograd and Cowan [1963]. However, in cooperative computation, the structure is usually configured so that computation will proceed to completion so long as the deleted subsystem is not an output system. To see this, let us return to the panel of physicians which we discussed (Section 4.1) as an analogy for the S-RETIC model of Kilmer and McCulloch. Consider the effect of removing one module/physician. In some happy cases, convergence on consensus may be sped up because the missing physician would have been grossly wrong in his initial diagnosis had he been present. In general, however, the absence of a member from a panel will slow convergence--the other physicians will still be able to reach their decision, but could certainly have profited from his input. Finally, the results may be disastrous when the patient's disease is so unusual that only a specialist has any chance of correct diagnosis, and it is that specialist who is missing.

This analysis makes it clear that, for the proper function of a system of cooperative computation, each subsystem must have a good enough model of the others to communicate effectively to get some ability to make up deletions. [This model can be a very crude one. For example, my model of the international banking system is simply that if I buy traveler's checks in one country, then I can spend the money in another country. Further information about the actual pattern of transactions required to bring this about is not going to affect my behavior.]

This last observation probably offers the key to the case against executive control in the implementation of the cooperative computation strategy. In executive control, we do not require subsystems to know anything about other subsystems--it is the job of the executive to switch in a subsystem as and when it is needed on the basis of executive monitoring of the completion of other tasks. However, if each system is truly complex, a central controller could be overloaded by simulating or in other ways studying each system in sufficient detail to determine the communication scheduling. While the throughput of local communication strategies may be suboptimal, the time lost in suboptimal computation may be far less than that required by an executive to actually compute the optimum.

With this, we should be a little bit more formal about what it is that lets one subsystem interact with another. I will use the term contract, introduced by Hewitt in his study of AI programming languages (see, e.g. Hewitt and Smith [1975]), although the basic idea is one that is as old as the general principle of hierarchical description: the idea that we can talk meaningfully about the interaction of high-level systems without having to continually spell out the intimate details of their working. A contract C_{ij} , then, specifies aspects of M_i 's operation sufficient for proper operation of M_j . M_i may be replaced by \hat{M}_i if

- (i) \hat{M}_i satisfies each C_{ij} and
- (ii) satisfaction of each C_{ji} suffices for proper operation of \hat{M}_i .

If either condition fails, iterative adjustment may be required. A basic evolutionary principle is that in making changes to satisfy old contracts, the system may develop the ability to satisfy new contracts.

We may thus suggest that our attempt to spell out more carefully what is involved in cooperative computation in the brain will not only tie in with what will become a dominant trend in computer science, but will also make contact with some of the most basic questions in evolutionary biology.

5.2 An Approach to Neurology

In this section we observe the way in which a number of neurological analyses of the effects of brain damage upon behavior are best viewed in terms of cooperative computation, rather than in terms of serial computation. For example, Luria has observed that a deep lesion of the premotor area yields elementary perseveration: a patient with this lesion, when instructed to draw a circle, will start to do so correctly, but will continue to trace around and around and around, without stopping spontaneously upon completion of the first circuit or two. Such a behavior does not seem well explained in terms of the alternating steps and tests of a serial program, but rather suggests that a subsystem for the control of movement is triggered by an initiator, while the completion of the plan is monitored by a separate subsystem (Figure 16)--and it is that 'completion monitor' that is damaged by the deep lesion of the premotor area.

As another example of cooperative computation in the motor system, we may recall the work of Orlovsky discussed in Section 1.3. There we saw (Figure 17) that the skeletal musculature is controlled not only by the spinal movement controller itself, but also by the brainstem adaptor which works in parallel to continually update the parametric adjustment of the basic movement controller.

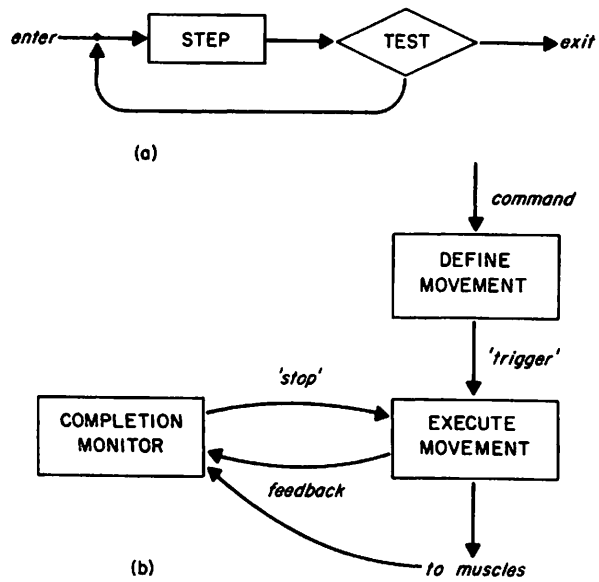


Figure 16:

A serial (a) and a cooperative computation (b) strategy for continuing a movement until completion.

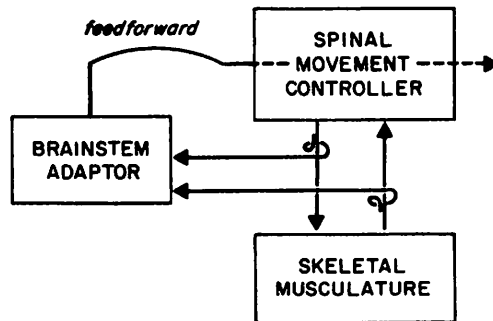


Figure 17:

The brainstem contains systems which act as an identification algorithm for the spinal control of locomotion.

This cooperative computation style is also appropriate for the approach to the frontal lobe taken by Nauta [1971]. A standard experiment for neuropsychologists is the delayed response test: a monkey is shown a food object being placed under one of two covers. He must then wait awhile before he is allowed to reach for the covers. A normal monkey can remember which cover has the object underneath for a long period of time; whereas the frontal monkey's performance soon degenerates to the random level. Many workers have thus interpreted this as a memory deficit, but for others, the interpretation of this as a "rigidity of central set" seems preferable: as if the context of the situation causes the response of the animal, rather than his memory of recent events.

Brenda Milner [1964] has studied some of the effects of frontal lobectomy in man. The patient is given a deck of 128 cards, with markings on each card differing in color, shape, or number of figures. The patient is to sort these into different piles--either on the basis of color or shape or number--inferring the criterion from observer feedback as to whether the placement of the cards is correct or not. Intriguingly, a frontal lobe patient will infer the feature, but will not change this inference when the observer switches the feature. Instead, the patient will simply note that the sorting does not seem to be proceeding very well any more. It is as if (Figure 18) the initial verbal instructions from the observer activate a system to establish the feature, but that once the feature is established, there is no internal subsystem--a 'success monitor'--which can reactivate the system for establishing the sorting criterion. In fact, loss of initiative is one of the most frequently remarked upon features of the frontal lobe patient.

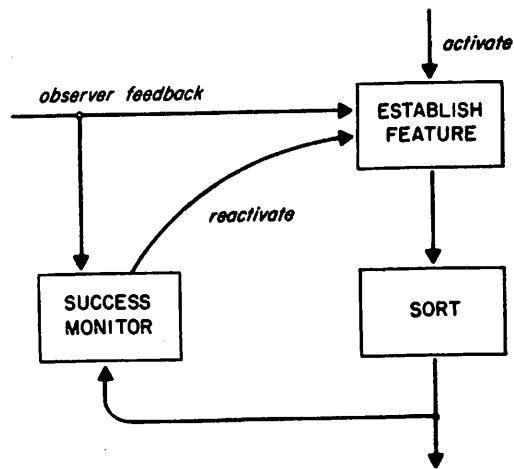


Figure 18:

A cooperative computation system for sorting cards on the basis of some feature specified by feedback from the observer.

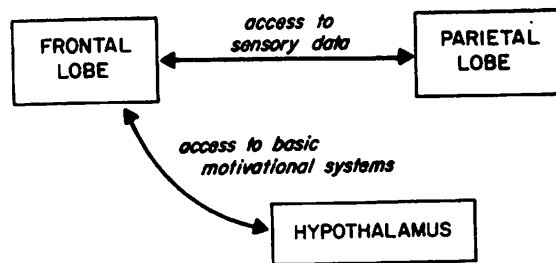


Figure 19:

The 'high road' and 'low road' to the frontal lobes.

Nauta has observed that the frontal lobe has extensive two-way connections not only to the parietal lobe--which contains an excellent 'map' of the position of the organism in its spatial world--but also to the hypothalamus, which contains the receptors and circuitry for the basic motivational systems of thirst, hunger, etc. (Figure 19). He suggests that this twofold connection of the frontal lobe allows for the continuous interleaving of the construction of high-level plans with the testing of 'how the plans taste', so that intellectual and basic drive criteria are continually inter-related.

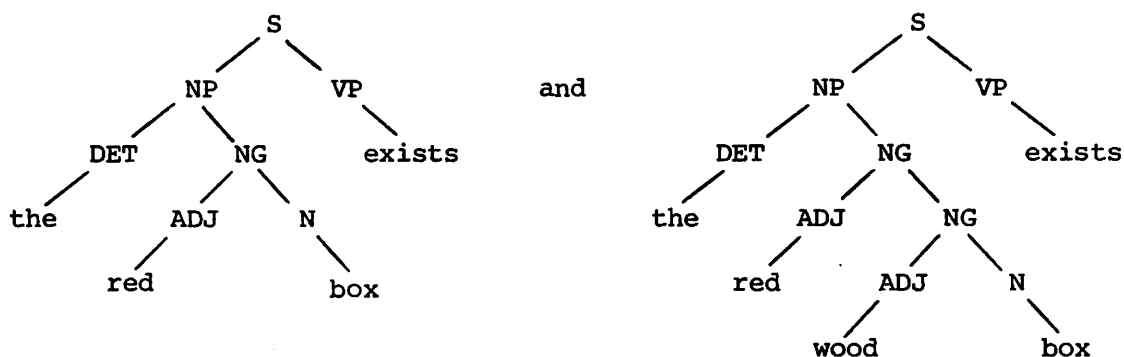
5.3 AI Approaches to Linguistics

This section is intended to suggest the difference between syntactic and semantic approaches to language analysis, to stress the role of question answering in computational semantics; and to suggest that the new complications brought about when a system must use speech rather than text for input demand a cooperative computation approach. This discussion of AI linguistics is by no means exhaustive, and the interested reader should turn to Schank and Colby [1973] and Schank and Nash-Webber [1975] for two excellent collections of contributions to this general field. However, it is an indication of the current state of the field of AI linguistics that these are collections of disparate approaches to the subject, devoid of any unifying theme. Wilks [1975] is one of the few attempts to compare different AI approaches to language.

The core of modern transformational linguistics (Chomsky [1965]) is the use of a phrase structure grammar to generate the deep structure of a sentence, and the use of transformations to manipulate that deep structure to get various surface structures. For example, given the rules

S → NP VP
 NP → DET NG
 NG → ADJ NG
 NG → N
 DET → the
 N → wood
 N → box
 ADJ → red
 ADJ → wood
 VP → exists

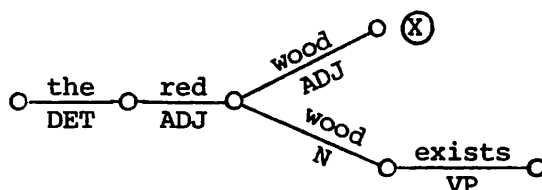
one can generate the derivation trees



which provide a syntactic analysis (analysis into grammatical constituents) of the two sentences "The red box exists," and "The red wood box exists," respectively. Transformations can then produce "The red wood box used to exist" (change of tense) and "The red wood box does not exist" (negation). [Of course, this glimpse of transformational grammar is superficial. For example, the derivation may yield a tree in which technical labels occur on some of the terminal nodes--and then certain obligatory transformations may have to be applied to yield a sentence of English.]

The starting point for a number of important AI analyses of language comes from Thorne, Bratley and Dewar [1968] who constructed a left to right parser for going from a sentence to its syntactic analysis. Briefly, the system works by moving from left to right, preserving at each stage all syntactic labels which can be given to the words of the string in a manner

consistent with the derivation rules of the grammar. Thus, for example, we would have two possible interpretations of the initial string "the red wood", but if the fourth word of the sentence is "exists" then it 'kills' the interpretation of "wood" as an adjective:



whereas if the fourth word is "box" it kills the interpretation of "wood" as a noun.

One of the interesting observations they made was that their system was able to work with a rather limited vocabulary of determiners, prepositions, and various key verbs. They found that with most sentences there was an unambiguous syntactic labelling of the constituents consistent with the labelling of the key words. They suggested that this was a mechanism by which one could learn the syntactic categories of new words, assigning them the only labelling which would allow the sentence in which they were first read or heard to be judged grammatical. The reader is of course familiar with this effect in such examples as Lewis Carroll's "Jabberwocky" in which we have no trouble in assigning syntactic categories to all the words of "Twas brillig and the slithy toves did gyre and gimble in the wabe."

Woods [1970] extended the approach of Thorne et al. to handle transformational grammars. Briefly, he noted that their approach could be formalized in terms of basic transition networks, in which an input word could trigger the passage from one node to the next precisely if the edge bore a syntactic label consistent with the word. Each node thus represented a classification of an input segment of the sentence. However, to handle

transformations, Woods found that it was not enough to have the basic transition network, but that one must have an augmented transition network in which registers were available at each node to maintain further information about the parsing to date.

With this background in syntax, we can now turn to semantics--to giving meaning to a sentence. The approach of many linguists has been what one might call 'minimal semantics', trying to find the least amount of information one must add to the syntactic categories of words to let a syntactic analysis go through. However, the goal of the AI linguist is to use language to communicate with a computer, and so the semantics are expressed in terms of actual performance of a robot or computer system. For example, Winograd's [1972] system is connected to a simulation of a hand-eye robot which can pick up and move blocks on a table top. The written inputs to the language system are either commands to manipulate the blocks, questions about the blocks world, or new information that the computer is to store about the blocks world. Given a command in natural English, the computer expresses its understanding by executing it. Given a question about the blocks world, the computer expresses its understanding by answering the question in English. Finally, given information about the blocks world, the computer will express its understanding by being able to make use of it when answering later questions to which it is relevant.

In the system described by Woods [1973], the computer has a data base on lunar rock samples, and it expresses its understanding of an English language question by generating the correct response. In both systems, the approach of the system is to use a basic syntactic analysis in terms of an augmented transition network to break the sentence into pieces which can then be reinterpreted to provide a program for carrying out the requests in

the input. Unlike the original Thorne system, these approaches do not try to keep all possible interpretations at any one time, but rather try to keep the most plausible interpretation, backing up to the last choice point when a given interpretation fails. Winograd makes use of semantic information to cut down the number of interpretations--for example, in the blocks world, there are no woods, only boxes, and so the interpretation of "wood" as a noun would be denied on semantic grounds.

It is beyond the scope of this paper to go into details of other currently fashionable approaches to AI linguistics. However, one should briefly mention the program of Schank to try and reduce sentences to a form structured from a few very basic primitives, using general properties of these primitives to derive inference about the meaning of a sentence, rather than using high-level information about the actual words. In other work--which we discussed briefly in Section 3.1--Schank and Abelson [1975] with their scripts are doing what Charniak [1975] and Minsky [1975] are doing with their demons and frames, namely tackling the problem of making available to a language understanding system the knowledge of the world with which we flesh out a very brief story. For example, consider the 'story': "John went into the restaurant and ordered a hamburger. He ate it with ketchup, paid the waiter, and left." How do we answer the question "Who brought him the hamburger?"? The answer, "the waiter", is based on our general knowledge of restaurants. Perhaps more interestingly, Schank notes that it is this knowledge that lets us accept the phrase "the waiter" in the above story rather than "a waiter", which is the form that we usually require to be used on the first appearance of a person in a story.

So far, we have assumed that the natural language entering the computer is typed in, so that there is no problem of character recognition.

A more interesting situation is that in which the human speaks to the computer. At first sight, it might seem that the problem breaks down into three parts: First: transform each word of spoken input into the standard computer representation of the word. Second: use syntax and semantics programs of the kind we have just described to build up an unambiguous representation of what each sentence really means. Third: answer the question, obey the commands, or store the information in the spoken input, whichever is appropriate.

However, this one-way flow of information does not work. People mumble, so that it is not clear which syllable they really spoke. People run one word into another, so that it is not clear where one word--or sentence--ends and another begins. People have different accents, so that one person's "yawl" is another person's "you all", and one person's "spine" is another's "Spain". Systems currently being built for speech understanding thus use a limited vocabulary and restrict the input to some microworld (such as the data base on lunar rock samples), using syntax, semantics and knowledge of the microworld to resolve ambiguities in the spoken word.

The first step is to turn the spoken input into a spectrogram--which shows how much energy there is in each frequency band. Experts can recognize some phonemes from the shapes of the bands in the spectrogram. This suggests that one use pattern recognition techniques to recognize the phonemes, and then chop the strings of phonemes up into words. The catch is that the decoding of spectrograms is highly error prone. An expert speech scientist, looking at one phoneme at a time, is only correct for about 30% of them if no context is used. In the other 70%, they might have anything from one to ten other guesses as to what the phoneme might actually be. However, if he uses a computer to

retrieve dictionary words which contain a sequence of phonemes meeting his list of partial specifications, the scientist can get 96% correct words, with most of the remaining errors confusing 'a' and 'the'. This study provides the basis for building a speech understanding system.

One such system, now in the pilot stage of development, is an outgrowth of the text understanding system of Woods [1973], LUNAR, which automated the answering of written queries that a geologist might ask about moon rocks, by translating sentences in ordinary English into one-shot computer programs for extracting the answer from the data base. The next stage in development of LUNAR is to program the computer so that it can also handle spoken queries. We now outline the way in which the LUNAR system is expected to handle speech when fully programmed.

Working on the spectrogram, the system might segment part of the input into three phonemes

$$\begin{pmatrix} \underline{l} \\ \text{or} \\ \underline{w} \end{pmatrix} \quad \begin{pmatrix} \text{front} \\ \text{vowel} \end{pmatrix} \quad \begin{pmatrix} \underline{s} \\ \text{or} \\ \underline{z} \end{pmatrix}$$

and consult a dictionary to find three words that meet these alternatives:

LESS

LIST

WAS

Given the type of questions it usually gets, LUNAR would order LIST as the most likely interpretation, followed by WAS and then LESS. The normal procedure, of which this is an example, is to use three phonemes (there may be phonemes between them) to direct each dictionary lookup. With only two phonemes, there may be too many consistent words; with more than three phonemes, errors may block retrieval of any dictionary entries.

Once a number of words have been found in this way, syntactic and semantic programs can be used to help work out 'blurred' phonemes. For example, if the system has decoded

LIST POTASSIUM RUBIDIUM RATIO

the syntactic procedure suggests that -s is the next phoneme, and so can give this hypothesis to the phoneme analyzer to test.

It turns out that short function words like FOR tend to be very fuzzily pronounced. Usually, the longer content words are much easier to pick out of the spectrogram, and the system can then use syntax to predict the small function words in a speech string, and then test phonemic match. This will usually be poor, and the next word is a good indicator of the likely correctness of the match. Thus the procedure of Thorne et al. of working from the function words is the opposite of the speech analysis approach, which works towards the function words.

The phonemic analysis often yields many distinct interpretations:

$$\begin{pmatrix} \text{BEEN} \\ \text{DID} \\ \text{DONE} \\ \text{ANY} \end{pmatrix} \begin{pmatrix} \text{EIGHTY} \\ \text{ANY} \end{pmatrix} \text{PEOPLE} \begin{pmatrix} \text{DONE} \\ \text{BULK} \end{pmatrix} \text{ANALYSES} \begin{pmatrix} \text{IN} \\ \text{ON} \end{pmatrix} \dots$$

and the system may use both syntax and knowledge of the rock sample micro-world to come up with the choice

HAVE ANY PEOPLE DONE ANALYSES ON . . .

calling a phonemic re-analysis to check the hypothesis that the first word is HAVE, rather than any of the alternatives previously listed.

In the Woods approach to speech understanding, there is much structuring done to ensure that only one aspect of the computation is running at any one time. By contrast, the Hearsay group (see for example Erman and Lesser [1975]) at Carnegie-Mellon University are explicitly looking at a

speech understanding system which is to be run on a network of PDP 11's. The idea is that each computer should contain a knowledge source. One source would be an expert at going from formants to phonemes, another expert at taking a string of words and making a syntactic analysis, while another can take partially analyzed sentences and look for semantic interpretations. The key idea of their system is that there should not be an executive controller which schedules tasks amongst these subsystems. Rather, there is a central communication center called a blackboard. Each knowledge source can take data from the level in which it is interested and work upon it, either returning to the blackboard an interpretation which wipes that question off the blackboard, or adding new questions. A basic level analyzer might take certain phonemes off the board, and return a number of consistent words. A syntactic analyzer might issue a request to check whether a particular word it requires is present at a particular place in the input. It is clear that this is very much in the spirit of cooperative computation in the sense in which we introduced Section 5.1--though unduly limited, perhaps, by the use of a single "blackboard". As the cost of microcomputers plunges well below \$1,000 one can expect an increasing interest in computer architecture based on networks of minicomputers and microcomputers. Thus the style--if not the particular details--of the Hearsay implementation, which is currently regarded by many AI workers as too unconventional, will, in my opinion, gain wide acceptance. It must be admitted that there are many problems attendant upon the current implementation. However, this is probably due to the novelty of implementing a computation on a minicomputer network, rather than to any inherent limitation in the approach itself. It thus seems safe to predict that cooperative computation will become ever more important in computer science, as well as providing an intriguing framework for the brain theorist.

6. EPILOGUE: UNITIES AND DIVERSITIES

My goal in this article has been to extract contributions from extant AI and BT literature, and meld them in an attempt to create a top-down brain theory, which has as its goal a coherent model of cooperative computation within which the computational roles of brain regions, and of neurons within those regions, can be analyzed. To the extent that I have succeeded, the reader may be puzzled by the inclusion of the word "diversities" in the title. In this brief concluding section, it is these diversities that reach center stage.

It must first be stressed that the literatures of AI and BT hardly overlap at all. While the problems of AI can be defined functionally, the problems of BT are defined structurally. Most AI workers are interested in making computers solve problems, plan, play games, understand natural language, etc. Virtually no eminent workers in AI feel it important to relate their research to actual brain mechanisms. Some, like Newell and Simon, are interested in using AI to shed new light on human problem solving, but such studies are psychological rather than neurophysiological. The brain theorist, on the other hand, studies neural networks, be they representative of actual brain structures, of a regular geometry conducive to simulation or the proving of theorems, or of interest mainly for the way in which their connections change over time. Most of the functions which dominate AI work are simply at too high a level for their expression in neural networks to be ripe for study. It is precisely to bridge this gap that I have urged the interest of a top-down approach to brain theory. AI techniques may help us understand how some overall behavior can be caricatured, at least, in terms of a scheme of cooperative computation by subsystems

whose function can be ascribed to regions of the brain. Bottom-up BT can then be used to analyze how the neural networks of each brain region might achieve the ascribed function. Of course, the neural modelling may well modify the hypothesis about the region's function, and this modification can be fed back to modify the top-down analysis.

Given the goal of this article, the reader will understand that the review of the literature has been highly selective, and that many truly excellent contributions to the fields of both AI and BT have been left out. In fact, whole fields have been omitted. For example, the study of learning has been excluded, though I shall attempt to do it some justice in Arbib [1976]. Here, it is enough to note the difference in emphasis, with AI looking at splicing and editing of programs to adapt them to a changing (or emerging) problem description, while BT stresses modification of synaptic weights to better adapt a neuron to its role as part of a pattern classification system, say.

We should next say something about the state of the art in both subjects. AI is by far the better established--for many years, SIGART (the ACM Special Interest Group of Artificial Intelligence) has published a Newsletter [address subscription enquiries to ACM Headquarters, 1133 Avenue of the Americas, New York, NY 10036], and 1975 saw the holding of the Fourth International Joint Conference on Artificial Intelligence. By contrast, 1975 marks the first year of publication of the Brain Theory Newsletter [address subscription enquires to The Center for Systems Neuroscience, GRC, University of Massachusetts, Amherst, MA 01002], and BT lacks any conference of its own, rating at best the occasional special session at conferences on neuroscience, on biomathematics, or on computer science or control theory. Both subjects lack any coherent body of theory accepted by the majority of

practitioners. The factionalism of AI is so great that, for example, workers at MIT--one of the leading AI research centers--rarely deign to publish their work in refereed journals, preferring to sell copies of their Ph.D. theses to other workers! BT is at a sufficiently immature stage that it is probably inappropriate to speak of factionalism--we simply haven't found each other yet! In any case, it is all too clear that both fields badly need a good dose of systematization. For too long, papers have been written without any attempt to build on the work of others, or to contribute to a unification of the achievements of several laboratories (cf. our remarks, for example, on AI linguistics in Section 5.3). Clearly, it is my belief--but not at all a widely held one--that this process of systematization can benefit from the type of melding process to which this article is a contribution.

This article points to the emergence of an increasing number of common principles for AI and BT. However, it would be wrong to infer from this that the two subjects will, or should, merge into one. AI must look from the common principles toward the world of computer science--to the writing of programs with increasing 'intelligence' to augment our use of computers by providing them with natural language understanding, automatically reconfigurable data bases, goal-oriented computer languages, etc. BT must look from the common principles toward the world of biology--to the increased understanding of brain regions and the neural networks which comprise them. Thus, no matter how far an AI and a BT worker travel together on solving a common problem, their paths must eventually diverge when the AI worker must find the most efficient data structure for currently available computers, and the brain theorist must reconcile his theoretical schemes with the reality of neural experiments. Nonetheless, it is my feeling that

much AI and BT now suffer from a premature casting of ideas into some programming language, or some neural network formalization; and that both fields will gain insight from a unified perspective.

REFERENCES

- G. I. Allen and N. Tsukahara [1974], "Cerebrocerebellar communication systems", Physiol. Rev., 54, 957-1006.
- 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 [1975a], "Segmentation, schemas, and cooperative computation", in Studies in Biomathematics (a volume in the MAA Studies in Mathematics) (S. Levin, ed.), in press.
- M. A. Arbib [1975b], "Parallelism, slides, schemas, and frames", paper presented at the NYU Symposium on Parallel Computation in Artificial Intelligence.
- 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. Händler, M. Spreng, eds.), R. Oldenbourg, 216-231.
- M. A. Arbib, G. F. Franklin and N. Nilsson [1968], "Some ideas on information processing in the cerebellum", in Neural Networks (E. R. Caianiello, ed.), Springer-Verlag, 43-58.
- H. B. Barlow, C. Blakemore and J. D. Pettigrew [1967], "The neural mechanism of binocular depth discrimination", J. Physiol., 193, 327-342.
- F. C. Bartlett [1932], Remembering, Cambridge University Press.
- C. C. Boylls [1975], "A Theory of cerebellar function with applications to locomotion. I. The physiological role of climbing fiber inputs in anterior lobe operation", COINS Technical Report 75C-6, Univ. of Mass.
- V. Braitenberg and N. Onesto [1960], "The Cerebellar Cortex as a Timing Organ," Congress, Inst. Medicina Cibernetica, Naples, 239-255.

- E. Charniak [1975], "Organization and inference in a frame-like system of common knowledge", in Schank and Nash-Webber, 42-51.
- N. Chomsky [1965], Aspects of the Theory of Syntax, MIT Press.
- P. Dev [1975], "Computer simulation of a dynamic visual perception model", Int. J. Man-Machine Studies, in press.
- R. L. Didday [1970], "The Simulation and Modelling of Distributed Information Processing in the Frog Visual System", Ph.D. Thesis, Stanford University.
- R. L. Didday and M. A. Arbib [1975], "Eye-movements and visual perception: A 'two-visual system' model", Int. J. Man-Machine Studies, 7.
- J. Doran and D. Michie [1966], "Experiments with the graph traverser program", Proc. Roy. Soc. Ser. A, 294, 235-259.
- J. C. Eccles [1973], "The cerebellum as a computer: Patterns in space and time", J. Physiol., 229, 1-32.
- L. Erman and V. Lesser [1975], "A multi-level organization for problem-solving using many, diverse, cooperating sources of knowledge", Proc. 4th Int. Joint Conf. Artificial Intelligence.
- R. E. Fikes, P. E. Hart and N. J. Nilsson [1972], "Learning and executing generalized robot plans", Artificial Intelligence, 3, 251-288.
- R. E. Fikes and N. J. Nilsson [1971], "STRIPS: A new approach to the application of theorem proving to problem solving", Artificial Intelligence, 189-208.
- H. G. Furth [1969], Piaget and Knowledge, Prentice-Hall.
- E. Goffman [1974], Frame Analysis: An Essay on the Organization of Experience, Harper Colophon Books.
- R. L. Gregory [1961], "The brain as an engineering problem", in Current Problems in Animal Behaviour (W. H. Thorpe and O. L. Zangwill, eds.), Cambridge University Press.
- S. Grillner [1975], "Locomotion in vertebrates: Central mechanisms and reflex interaction", Physiol. Rev., 55, 247-304.
- 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.
- P. Hart, N. J. Nilsson and B. Raphael [1968], "A formal basis for the heuristic determination of minimum cost paths", IEEE Trans. Syst. Sci. Cybernetics, SSC-4, 100-107.
- C. Hewitt [1969], "Planner: A language for proving theorems in robots", in Proceedings of the International Joint Conference on Artificial Intelligence (D. Walker and L. Norton, eds.)

- C. Hewitt and B. Smith [1975], "Towards a programming apprentice", IEEE Trans. Software Engineering, 1, 26-45.
- G. Holmes [1939], "The cerebellum of Man", Brain, 62, 1-30.
- D. H. Hubel and T. N. Wiesel [1962], "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", J. Neurophysiol., 28, 229-289.
- D. Ingle [1968], "Visual releasers of prey-catching behavior in frogs and toads", Brain Behav. Evol., 1, 500-518.
- D. Ingle and J. M. Sprague [1975], "Sensorimotor function of the midbrain tectum", Neuroscience Research Program Bulletin, 13, No. 2.
- M. Ito [1974], "The control mechanisms of cerebellar motor systems", in The Neurosciences: Third Study Program (F. O. Schmitt and F. G. Worden, eds.), MIT Press, 293-303.
- B. Julesz [1970], Foundations of Cyclopean Perception, Chicago University 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.
- K. S. Lashley [1929], Brain Mechanisms and Intelligence, University of Chicago Press.
- K. S. Lashley [1951], "The problem of serial order in behavior", in Cerebral Mechanisms in Behavior (L. Jeffress, ed.), Wiley-Interscience, 112-136.
- J. Y. Lettvin, H. Maturana, W. S. McCulloch and W. H. Pitts [1959], "What the frog's eye tells the frog's brain", Proc. IRE, 47, 1940-1951.
- J. Lighthill, et al. [1973], Artificial Intelligence: A paper symposium, Science Research Council.
- R. R. Llinas [1975], "The cortex of the cerebellum", Sci. Amer. January, 232, No. 2, 56-71.
- 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.
- G. A. Miller, E. Galanter and K. H. Pribram [1960], Plans and the Structures of Behavior, Holt, Rinehart and Winston.
- B. Milner [1964], "Some effects of frontal lobectomy in man", in The Frontal Granular Cortex and Behavior (J. M. Warren and K. Akert, eds.), McGraw-Hill, 313-334.
- M. L. Minsky [1975], "A framework for representing knowledge", in The Psychology of Computer Vision (P. H. Winston, ed.), McGraw-Hill, 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.
- J. I. Nelson [1975], "Globality and stereoscopic fusion in binocular vision", J. Theor. Biol.
- A. Newell, J. C. Shaw and H. A. Simon [1960], "Report on a general problem-solving program for a computer", Information Processing: Proc. Int. Conf. Information Processing, UNESCO, 256-264.
- A. Newell and H. A. Simon [1972], Human Problem Solving, Prentice-Hall.
- G. N. Orlovsky [1972], "The effect of different descending systems on flexor and extensor activity during locomotion", Brain Res., 40, 359-371.
- J. D. Pettigrew, T. Nikara and 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], "Scene labelling by relaxation operations", TR-379, Computer Science Center, University of Maryland, College Park.
- E. D. Sacerdoti [1974], "Planning in a hierarchy of abstraction spaces", Artificial Intelligence, 5, 115-135.
- R. C. Schank and R. P. Abelson [1975], "Scripts, plans and knowledge", Proc. Fourth International Joint Conference on Artificial Intelligence.
- R. C. Schank and K. M. Colby (eds.) [1973], Computer Models of Thought and Language, W. H. Freeman.
- R. C. Schank and B. L. Nash-Webber [1975], Proc. Workshop on Theoretical Issues in Natural Language Processing, Association for Computational Linguistics.
- R. C. Schank and C. J. Reiger, III [1974], "Inference and the computer understanding of natural language", Artificial Intelligence, 5, 373-412.
- R. C. Schank and Y. Wilks [1974], "The goals of linguistic theory revisited", Lingua, 34, 301-326.
- M. E. Scheibel and A. B. Scheibel [1968], "The brainstem core--An integrative matrix", in Systems Theory and Biology (M. Mesarovic, ed.), Springer-Verlag, 261-285.

- C. F. Schmidt [1975], "Understanding Human Action", Proc. Workshop on Theoretical Issues in Natural Language Processing, 196-200.
- J. R. Searle [1965], "What is a speech act?", Philosophy in America (Max Black, ed.), Allen & Unwin, 221-239.
- O. G. Selfridge [1959], "Pandemonium: A paradigm for learning", Mechanization of Thought Processes, London, H.M.S.O., 513-526.
- R. B. Stein [1974], "Peripheral control of movement", Physiol. Rev., 54, 215-243.
- J. Szentágothai and M. A. Arbib [1974], Conceptual Models of Neural Organization, NRP Bulletin 12, No. 3. (Republished by MIT Press, 1975.)
- J. P. Thorne, R. Bratley and H. Dewar [1968], "The syntactic analysis of English by machine", in Machine Intelligence, 3, (D. Michie, ed.), Univ. of Edinburgh Press, 281-309.
- Y. Wilks [1974], "Natural language understanding systems within the AI paradigm", Stanford Artificial Intelligence Lab, Memo, AIM-237.
- S. Winograd and J. D. Cowan [1963], Reliable Computation in the Presence of Noise, MIT Press.
- T. Winograd [1972], Understanding Natural Language, Academic Press.
- W. A. Woods [1970], "Transition network grammars for natural language analysis", Comm. ACM, 10, 591-606.
- W. A. Woods [1973], "Progress in natural language understanding--An application to lunar geology", Proc. Nat. Comp. Conf., 441-450.
- R. H. Wurtz and M. E. Goldberg [1971], "Superior colliculus cell responses related to eye movements in awake monkeys", 171, 82-84.