Modularity and Interaction of Brain Regions Underlying Visuomotor Coordination*

Michael A. Arbib

Center for Systems Neuroscience
Computer & Information Science Department
University of Massachusetts
Amherst, Massachusetts 01003

COINS Technical Report 85-29

---

*Preparation of this paper was supported in part by NIH NS14971.

Modularity and Interaction of Brain Regions Underlying Visuomotor Coordination *

Michael A. Arbib

Center for Systems Neuroscience

Department of Computer and Information Science

University of Massachusetts

Amherst, Massachusetts 01003

---

* Preparation of this paper was supported in part by NIH NS14971.

1. Modules in the Brain

It is a standard notion that a complex system may be analyzed by being decomposed into a set of interacting subsystems. Such a decomposition succeeds insofar as we can understand the relation between the inputs and outputs of each individual subsystem, and that the interactions between the subsystems can be explained via suitable connections between various of their inputs and outputs, without further analysis of variables internal to the subsystems. Such a decomposition is structural to the extent that the subsystems can be mapped onto physical substructures of a physical structure embodying the overall system. In this section, I show that neuroscientists have long sought structural decompositions of the brain, and in some cases referred to the physical substructures as modules. Recently, Fodor has popularized the use of the term "module" to denote a unit in a functional decomposition of a cognitive system, but a subsystem that meets constraints beyond those specified above. I shall argue that Fodor's analysis of cognitive systems is flawed and that the restrictions he introduces are not useful. Consequently, I shall use the term "module" as a synonym for the term "subsystem" defined above.

The work of the nineteenth century neurologists led us to think of the brain in terms of large interacting regions each with a more or less specified function, and this localization was reinforced by the work of the anatomists at the turn of the century who were able to subdivide the cerebral cortex on the basis of cell characteristics, cytoarchitectonics. It was at this same time that the discoveries of the neuroanatomist Ramon y Cajal and the neurophysiologist Sherrington helped establish the neuron doctrine, leading us to view the function of the brain in terms of the interaction of discrete units, the neurons. The issue for the brain theorist, then, is to map complex functions, behaviors, patterns of thought, upon the interactions of these rather large entities, anatomically defined brain regions, or these very small and numerous components, the neurons. This has led many neuroscientists to look for structures intermediate in size and complexity between brain regions and neurons to provide stepping stones in an analysis of how neural structures subserve various functions. One early example was the Scheibels' [1958] suggestion that the reticular formation could be approximated by a stack of "poker chips" each incorporating a large number of neurons receiving roughly the same input and providing roughly the same output to their environments. This modular decomposition of the reticular formation provided the basis for the Kilmer and McCulloch model, RETIC (Kilmer, McCulloch and Blum [1969]). In another direction, the theoretical ideas of Pitts and McCulloch [1947] combined with the empirical observations of Lettvin, Maturana, McCulloch and Pitts [1959] on the frog visual system to suggest that one might think of important portions of the brain in terms of interacting layers of neurons, with each layer being retinotopic in that the position of neurons in the layer was correlated with position on the retina, and thus in the visual field. Powell and Mountcastle [1959] working in somatosensory cortex, followed by Hubel and Wiesel [1974] working in visual cortex, established the notion of the column as a "vertical" aggregate of cells in visual cortex, again working on a common set of inputs to provide a well-defined set of outputs. With all these considerations, the notion of the brain as an interconnected set of modules -- intermediate in complexity between neurons and brain regions -- was well established. For example, Szentagothai and Arbib [1974], in their monograph on *Conceptual Models of Neural Organization*, wrote that

> The concept of a modular structure or arrangement of the neuropil has two basic sources:
>
> (1) a more indirect one from the notion that neuronal networks ought to be subdivided into distinct functional units; and
>
> (2) a direct one from observation of the neuropil...Recent anatomical data on the cerebral cortex suggest the existence of both a fine grain and coarse grain of modular organization.

Subsequently, Mountcastle [1978], in an essay titled "An Organizing Principle for Cerebral Function: The Unit Module and the Distributed System", wrote that

> The large entities of the brain we know as areas (or nuclei) ... are themselves composed of replicated...modules...Each module processes information from its input to its output and in that processing imposes transforms determined by the general properties of the entity and its extrinsic connections. ... Closely linked ... subsets of modules ... form precisely connected but distributed systems. ... A single module of an entity may be a member of several (but not many) such systems.

With this, it is clear that the concept of a "module" is well established within neuroscience as a structural entity, but our task in this paper is to confront it with the notion of a module as a functional entity, as developed, for example, in the elegant monograph The Modularity of Mind by Jerry Fodor. To see this, it will help to distinguish what has been called top-down brain theory from bottom-up theory, and see how they are brought together in what might be called "middle-out" brain theory. Top-down theory is essentially functional in nature, in that it starts with the isolation of some overall function, such as some pattern of behavior or linguistic performance or type of perception, and seeks to explain it by decomposing it into the interaction of a number of subsystems. What makes this exercise brain theory as distinct from cognitive psychology is that the choice of subsystems is biased in part by what we know about the function of different parts of the brain, as obtained for example by analysis of the effects of brain lesions, so that there is some attempt to map the subsystems onto anatomical regions.

In bottom-up brain theory, the emphasis tends to be on neural circuits. Given a specific set of neurons, the attempt is to use them to implement a given function, or to analyze given circuitry to determine what functions it can perform. Clearly, the primary sources of data for such bottom-up brain theory come from both neuroanatomy -- in looking at the detailed interconnections of neurons -- and from neurophysiology -- in studying the behavior of the network under varying conditions.

In its full development, brain theory incorporates a cycle of both top-down and bottom-up modelling. An attempt is made to map functional units onto brain regions, constrained for example by lesion studies. But this map is not one to one, and a further constraint is to try to implement regional functions via neural networks meeting anatomical and physiological constraints. New information may not only yield different suggestions as to how circuitry may subserve a given function, but -- as we shall see well exemplified in section 4 -- may lead to changes in our ideas about how functions are distributed around the regions of the brain. In particular, we shall see that there is a continual tension between functional decomposition and structural decomposition, and that in general a given "functional module" may be subserved by the interaction of several "brain modules", and that a given "brain module" may be involved in subserving a number of different functions -- as was indeed suggested by the quote from Mountcastle above.

2. Fodor's Taxonomy of the Mind

For Fodor, a computational theory of cognitive science seeks a set of mechanisms each of which provides a characteristic pattern of transformations of mental representations [F13]. ** For him, a computational process is by definition syntactic [F40]. Motivated by Gall's view of faculties, Fodor seeks cognitive mechanisms which form modules in the sense that they are [F21, F37]

(i) Domain-specific;

(ii) Innately specified;

(iii) Associated with distinct neural structures; and

(iv) Computationally autonomous.

In the present paper, I shall not address claims (ii) and (iii), but shall restrict myself to setting forth my argument that cognitive science will not be served by an emphasis on modules which are domain-specific and computationally autonomous.

The term "domain-specific" seems to refer to gross modalities like 'vision' or 'language' rather than Gall's faculties or the "domains" or "microworlds" of current AI research. The key concept is that of computational autonomy: modules "do not share, and hence do not compete for, such horizontal resources as memory, attention,

---

** A reference like [F13] refers to page 13 of Fodor's The Modularity of Mind, while [FIII.5] refers to Section III.5 of that volume.

judgement, intelligence, etc." [F21]. More generally, the modules considered by Fodor are informationally encapsulated [FIII.5], which means that, while there may be internal feedback between the representations within the module, these internal representations are not involved in paths to or from external modules. Moreover, Fodor argues that there is no external feedback path whereby the output of the module can affect its input (Figure 1).

Fodor argues for an exclusive but not exhaustive functional taxonomy of the mind as divided into transducers, input systems, and central processors [F41]. He views the input systems as delivering representations that are most naturally interpreted as characterizing the arrangement of things in the world [F42], and then lumps perceptual systems with language as constituting the input systems, with the note [F45] that what underwrites the correspondence between, say, visual stimuli and distal layouts are (roughly) the laws of light reflectance, whereas what underlies the correspondence between token utterances and distal layouts is (roughly) a convention of truth-telling which makes it possible to infer from what one hears to the way that the world is.

Fodor's central hypothesis about the mind is that the input systems (perceptual systems and language) are precisely the modules of the mind, "domain-specific computational systems characterized by informational encapsulation, high speed, restricted access, and neural specificity [F101]". "The input systems being informationally encapsulated compute representations of the distal layout on the basis of less information about the distant layout than the organism has available. Such representations want correction in the light of backgroud knowledge and of the simultaneous results of analysis in other domains. Call the process of arriving at such corrected representations 'the fixation of preceptual belief' [F102]".

Fodor then argues [F103] that an interface between perception and utilities must take place <u>somewhere</u> if we are to use the information that input systems deliver in order to determine how we ought to act. Fodor then argues that these central processes for thought and problem solving are unencapsulated and so are not plausibly viewed as modular [F103]. Although Fodor gives no diagrams in his book, his theory of the mind is, I think fairly, captured in Figure 2.

Some comments are in order. First, we reiterate that Fodor's use of "domain" seems to refer to "language or a sensory modality". Second, Fodor gives no analysis of action and motor control, so it is open as to whether or not he would regard these as modular, e.g., with modules for locomotion, manipulation, and speech. But third, and this provides the starting point for our critique, Fodor is silent about which transducers feed the language module. Ignoring braille, I have arbitrarily shown auditory and visual transducers providing this input. But once we insist that, e.g., the visual transducers deliver input to vision, audition and language without cross-talk between these modules (for such cross-talk would

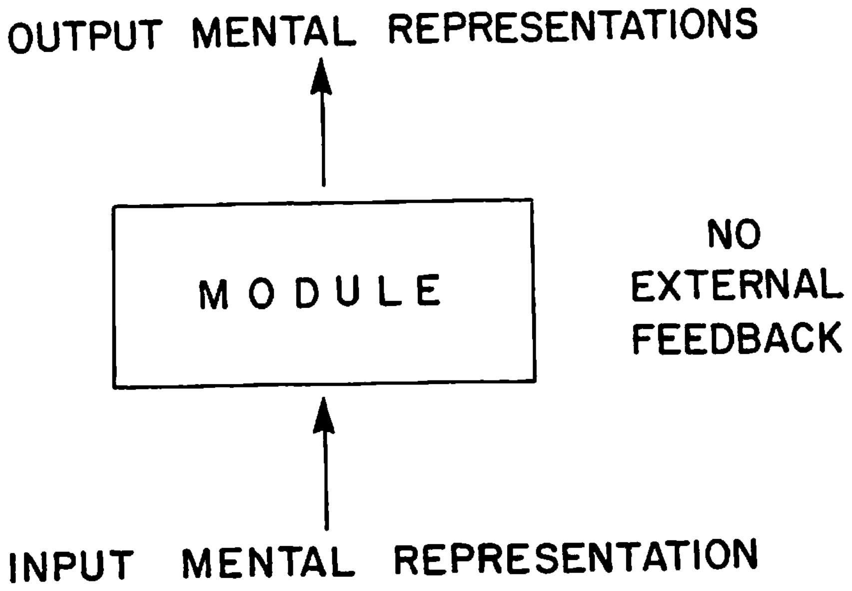


Figure 1. A key property of a module is that it be computationally autonomous and informationally encapsulated. Moreover, its input cannot depend upon its output via paths internal to the cognitive system in which it is embedded.

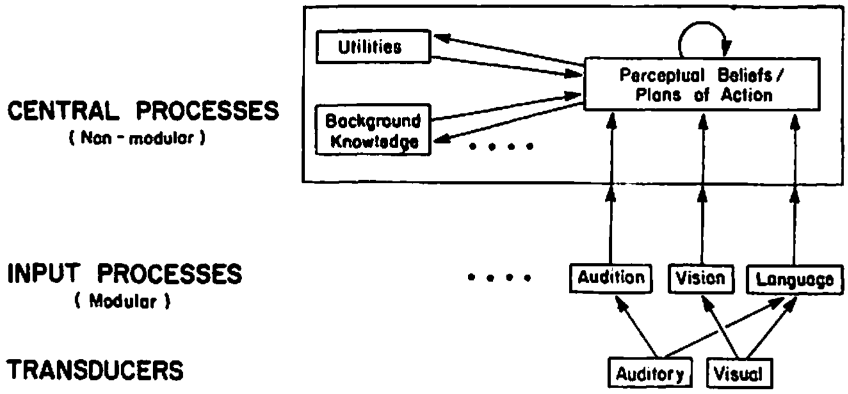


Figure 2. Fodor's Taxonomy of the Mind. The arrows linking the central processes have no significance other than to indicate the promiscuity of data-flow which renders the central processes non-modular. By contrast, the lack of feedback paths around each input process is crucial to Fodor's theory.

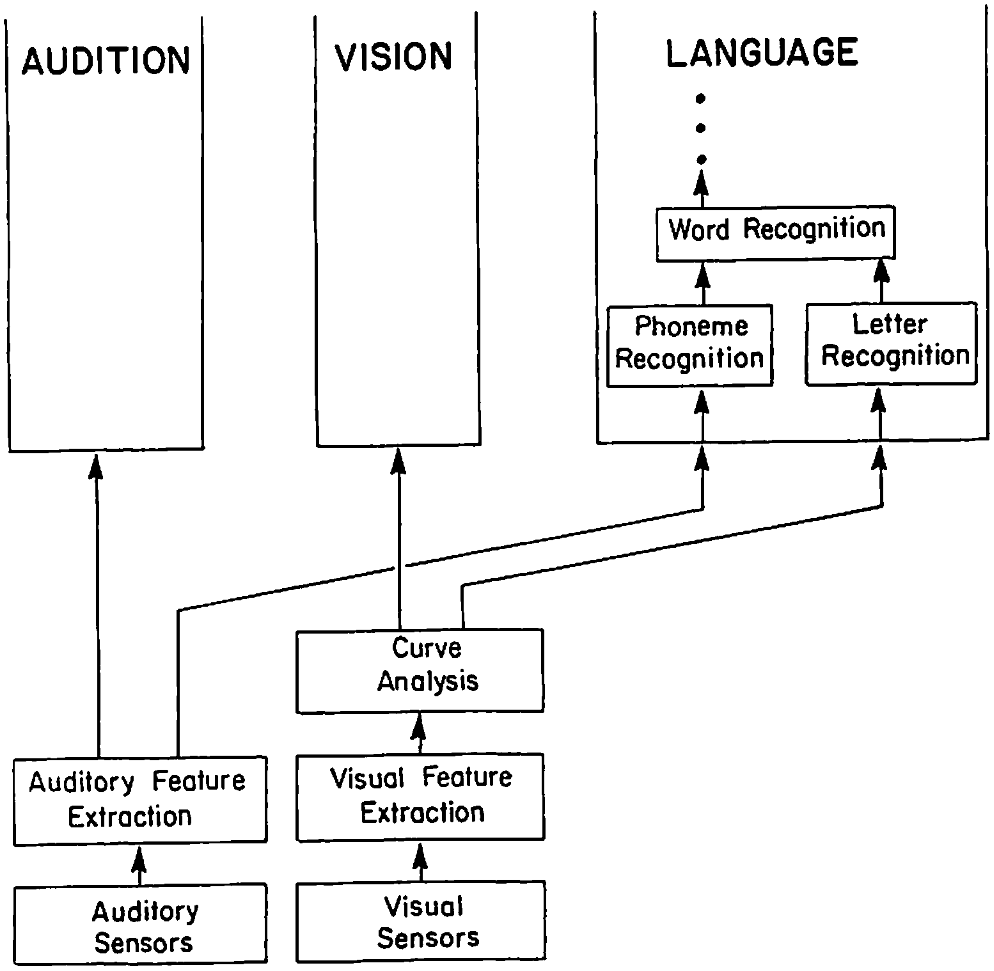


Figure 3. A hypothetical refinement of the transducers and input processes of Figure 2.

break the posited informational encapsulation), the transducers no longer "merely" transduce sensory signals, but carry out some moderately sophisticated transformations of their own, as sketched in Figure 3. For example, since the visual recognition of curves is helpful both in seeing objects and in recognizing letters, we must either make this curve analysis part of the visual transducer or we must see it as constituting an input module which is not domain-specific. In either case, the distinction between transducers and input processes now seems less comfortable, and will become more so if we try to refine the language, vision, and touch modules in a way that will accommodate the learning of Braille and sign language, for now the intermixing of language with touch and motor control makes claims for the informational encapsulation of the modules of Figure 2 well-nigh untenable.

However, my fundamental point is that Fodor's modules are <u>too large</u>. It is clear from Figure 3 that a computational theory of cognition must use a far finer grain of analysis than that offered by Fodor (Figure 2). Fodor offers <u>big</u> modules, argues vociferously that they are computationally autonomous, and despairs at the problem of explaining the central processes, since they are not informationally encapsulated (more of this in section 6). By contrast, my approach is to analyze the brain in terms of smaller components. Since the interactions between these components play a vital role in my models, the case for autonomy of large modules becomes less plausible -- a useful measure of parsimony, rather than a fundamental principle.

In the next three sections I shall try to exemplify this approach to the study of modularity and interaction of brain regions by looking at a number of models of visual perception and of visuomotor coordination. In the first two and one half of these sections, I will be guided by the slogan set forth in The Metaphorical Brain [Arbib 1972] that "The brain should be modelled in terms of distributed action-oriented computation in layered somatotopically organized machines." The term "distributed" emphasizes the notion that the brain is made up of many different systems which are simultaneously active, and so is not to be modelled in terms of serial computation, in which one localized operation is conducted at a time. Secondly, the notion that the brain is "action-oriented" emphasizes that we should not think of, for example, vision purely in terms of rendering some sort of objective representation of the visual world within the brain, but rather should ask how that visual representation can provide information that is relevant to the activity of that particular organism. For example, we shall see that the visual system of a frog is quite different from the visual system of a human, even though we shall at the same time strive to find general principles which will help us understand what is involved in the structure of these systems. Finally, the slogan that the brain is "a layered somatotopically organized machine" will correspond to the claim, made in section 1, that many parts of the brain can be analyzed in terms of the interaction of layers of similar components. However, in the last part of my discussion of vision in section 5, I will point out that there are processes that do not seem to be naturally thought of in terms of somatotopic or retinotopic computation in a layered structure. This will serve to introduce the notion of <u>schema</u> as a fine-grained functional module, corresponding for example to the knowledge required to recognize a house or to grasp a mug. In this paper, I shall offer no data on the extent to which schemas may be considered as psychologically or neurophysiologically verifiable, but simply show their computational role in a model of high-level vision. In other writings, I have presented the notion in the context of general visuomotor coordination [Arbib 1981], developed its use for explaining skilled manual behavior [1985], and explored with my colleagues the use of schemas in developing various models of language [Arbib, Conklin and Hill, in press].

3. Depth Perception

A chapter in brain theory usually starts by trying to characterize some overall function which one thinks might constitute a "brain module" in that one might hope to find specific brain circuitry devoted to its implementation. For example, it seems reasonable to think that "solving differential equations" would not be a suitable module for such a study, and that our eventual understanding of its brain implementation will come by seeing how cerebral circuitry capable of implementing many different human skills can be tuned by experience and instruction to subserve this particular mathematical ability. By contrast, we have every reason to think of vision as having special circuitry, from the retina of the eye to the tectum of the midbrain and the lateral geniculate nucleus of the thalamus up to a number of regions designated as visual cortex. In fact, the work of the last twenty years has taken us from beyond this conception, and we now talk of the "many visual systems" finding that there are now tens of anatomically distinct brain regions each subserving distinctive functions of vision. Within this context, then, we may seek to find natural functional modules which subserve part of the task of vision, thus focusing our attention on a more restricted system of neural circuitry. One such natural candidate is _depth perception_ -- which enables us to determine the world in terms of objects located at various distances from us. This is clearly very important from an action-oriented perspective, since the way in which we will interact with objects will depend crucially on how near or far they may be.

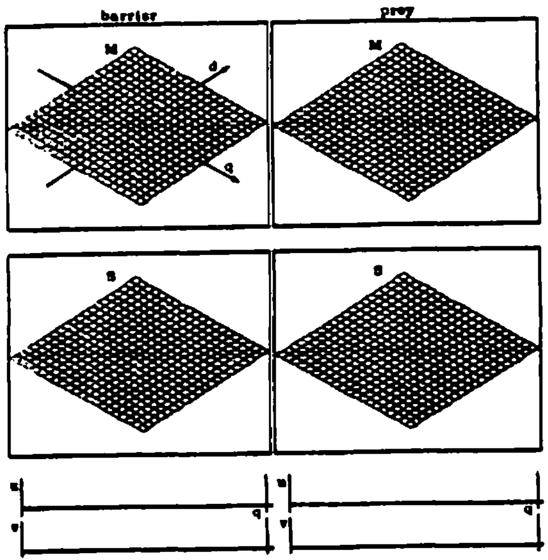
It is a familiar experience from using stereo viewers that the view of a three dimensional scene presented at the left eye differs from that presented to the right eye, and that the disparity or displacement between these two images provides the crucial cue as to the distance in space from which they come. A key concern from the nineteenth century was whether depth perception comes before or after pattern recognition. Is it that the brain takes the image from each eye separately to recognize, for example, a house therein, and then uses the disparity between the two house images to recognize the depth of the house in space; or is it that our visual system matches local stimuli presented to both eyes, thus building up a depth map of surfaces and small objects in space which provides the input for perceptual recognition? It was the great achievement of Bela Julesz to invent the method of random dot sterograms, providing stereo pairs each of which contains only visual noise, but so designed that the visual noise was correlated. Patches of random light and dark presented to one retina were identical to, but at varying disparities from, patches of light and dark presented to the other retina. Julesz found that people were in fact able to carry out the appropriate matching to see surfaces stippled with random patterning at varying depths in space. In other words, without precluding that some depth perception could follow pattern recognition, he did establish that the forming of a depth map of space could precede the recognition of pattern. He offered a model of this process in terms of cooperative computation involving a somewhat Heath Robinson array of magnetic dipoles connected by springs.

For the brain theorist the issue was thus raised: "could the depth map be computed by a cooperative process involving realistic neurons?", and the data on what constituted realistic neurons was provided by the work of Barlow and various co-workers. One of the first papers to address this issue was that by Arbib, Boylls and Dev [1974], who built a neural net cooperative computation model for building the depth map "guided by the plausible hypothesis that our visual world is made up of relatively few connected regions". The neural manifold of this model had cells whose firing level represented a degree of confidence that a point was located at a corresponding position in three dimensional space. The neurons were so connected via inhibitory interneurons as to embody the principle that cells which coded for nearby direction in space and similar depth should excite each other, whereas cells which corresponded to nearby direction in space and dissimilar depth should inhibit each other. It was shown by computer simulation by Dev [1975] and later established by mathematical analysis by Amari and Arbib [1977] that this system did indeed yield a segregation of the visual input into connected regions. Later, a variant of this model was published by Marr and Poggio [1976], and in subsequent writings Marr took the "plausible hypothesis that our visual world is made up of relatively few connected regions" and showed how it could be developed into an elegant mathematical theorem relating the structure of a depth perception algorithm to the nature of surfaces in the physical world.

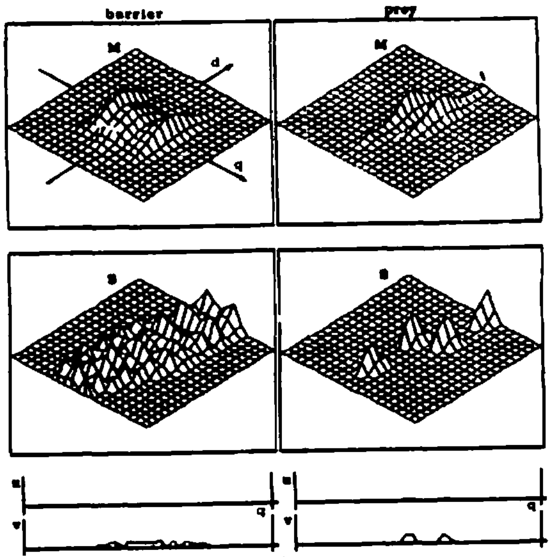
With this work, then, it was established that the perception of depth maps could be constructed by a method of computation that was guided by the hypothesis that the world was made up of surfaces, and that the algorithm could involve some form of cooperative computation. However, the cooperative computation algorithms discussed above exhibited the problem of false minima. Consider, for example, a picket fence. Suppose by pure randomness that the system starts by matching a number of fence posts presented to one eye with the images of their neighbors one to the left presented to the other eye. In the cooperative computation model, this initial mismatch could co-opt the possible choices of neighbors, and end up with a high confidence estimate that the fence was at a different depth from that at which it actually occurred. This provides a local "energy minimum" for the algorithm. The question then arises of how one could come up with an algorithm that would avoid some at least of these false minima. The answer provided by Marr and Poggio [1979] can be seen as rooted in two contributions, one from machine vision, and one from psychophysics. Within the machine vision community, such workers as Rosenfeld and Hanson and Riseman had put forward the idea of pyramids or processing cones; the notion that one could look at a visual image at different scales of resolution, and that for some problems a blurred image would allow one to quickly extract a first approximation to needed information about the image, information that would in fact be costly and time consuming to extract when working at the full detail of the original image. Meanwhile, the Cambridge school of psychophysists headed by Fergus Campbell had discovered that the brain itself appeared to employ a form of multiple levels of resolution -- that there were cells tuned to different spatial frequencies, and that these spatial frequencies tended to fall into four or perhaps five different channels. This led Marr and Poggio to develop a system in which, with hardly any cooperative computation, a fairly confident rough depth estimate for different surfaces could be made using the low spatial frequency channels, and then a more detailed spatial map could be sculpted on the first approximation through the more detailed disparity information provided via channels of higher spatial frequency. Subsequent psychophysical studies by Frisby and Mayhew have shown that there is much to be said for this model as a model of human depth perception, although they have provided data which refined the model, and show that some measure of cooperativity is required. Prazdny has come up with further material for modifying and changing the model.

However, it must be noted that the fact that a brain mechanism is employed to implement a particular functional module by the brain of one animal does not imply that this is the mechanism used by a different species. It is known that frog and toad can snap with moderate accuracy at prey located in the monocular visual field, and this led Ingle to hypothesize that for the frog it was accommodation, focal length information for the lens, that subserved depth perception. More detailed experiments by Collett, involving placing on the nose of the toad spectacles which used either prisms or lenses, showed that the story was more complex. Collett was able to show that the monocular animal did indeed use accommodation as the depth cue, but that in an animal with prey in the binocular field, the major depth cue was disparity, with accommodation cues exerting perhaps a six percent bias on the depth judgement based on disparity. This led Donald House [1982] in my laboratory to suggest that another route to solving the problem of false minima was to use accommodation cues to bias disparity. This is exemplified by Figure 4 in which we see two depth maps corresponding to two "worms" presented as visual targets to the simulated toad. In the accommodation map, the level of activity at a particular position and depth corresponds to how sharp an image was obtained at that position when the lens was focused at the given depth. We thus see that the activity has one peak for each worm, but that the peak is rather broad, giving poor localization information. By contrast, the disparity map of depth gained by pairing stimuli on the two eyes has the problem that, although it gives precise localization of the two worms, it also gives precise localization of the "ghost worms" got by matching a stimulus on one eye with the wrong stimulus on the other eye. The key observation here is that the two sources of depth information provide complementary information. House's model uses a variant of the Dev model to refine the depth estimates within each depth field, but adds mutual coupling between the models so that activity localized within one map helps increase activity at the corresponding locus on the other map. Simply looking at the figure, without knowing any further details, it is clear that this interaction can yield both a sharpening of the peaks in the accommodation map and a suppression of the "ghost peaks" in the disparity map, finally converging on a state in which both maps agree and present a sharp accurate localization of the targets.

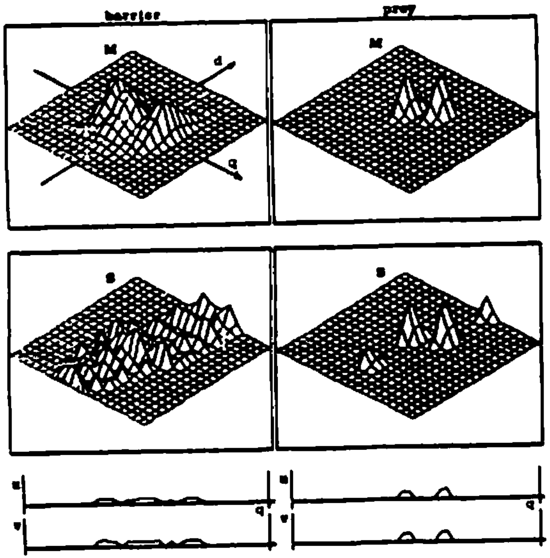
In concluding this section, we make several points. Firstly, that there is no unique algorithm for solving a given problem, in part because many different sources of information can be employed. Secondly, any one source of information will be incomplete, and it will often take the skillful deployment of several sources of information -- whether it be disparity information at several levels of resolution, or disparity and accommodation information in conjunction -- to yield a far better estimate than could be gained by using one source alone. And thirdly, we have seen how these considerations lead us to a far finer modular decomposition than that offered by Fodor (and which we have diagrammed in Figure 3).


barrier
prey
M
d
q
M
S
S
u
q
v
u
q
v

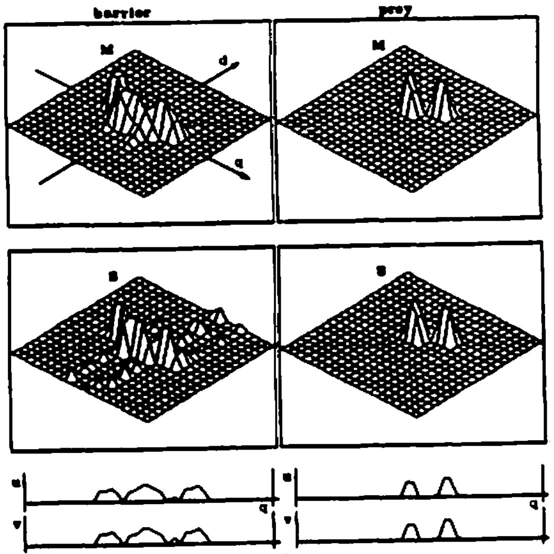
a


barrier
prey
M
d
q
M
S
S
u
q
v
u
q
v

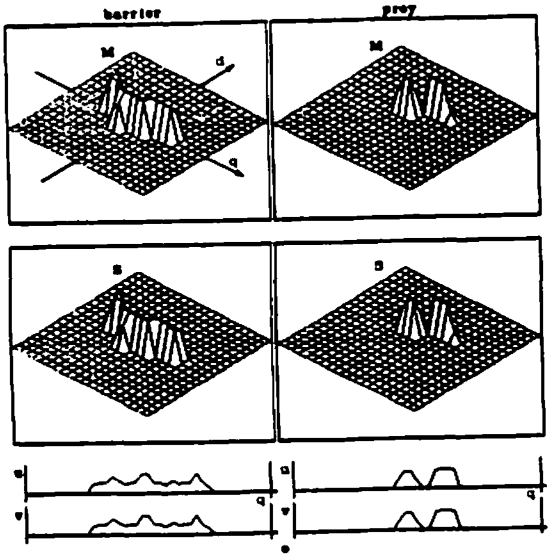
b


barrier
prey
M
d
q
M
S
S
u
q
v
u
q
v

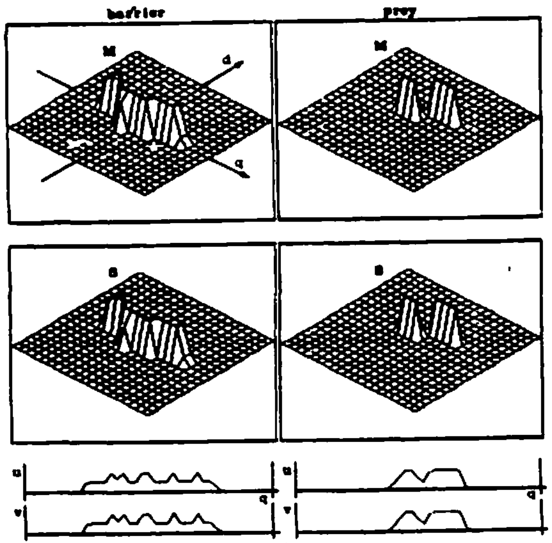
c


barrier
prey
M
d
q
M
S
S
u
q
v
u
q
v

d


barrier
prey
M
d
q
M
S
S
u
q
v
u
q
v

e


barrier
prey
M
d
q
M
S
S
u
q
v
u
q
v

f

Figure 4. Time Course of the Model -- Base Case: The time course of the depth model from its initially-inert state (a) to a satisfactory depth segmentation (f) is shown here. The full monocular/binocular model was used with two input planes: one for accommodation and one for disparity. All figures are in the retinal angle vs. disparity coordinate system. Successive figures are temporally spaced 1-1/3 field time-constants apart. Thus, the elapsed simulation time represented is nearly 7 time-constants. The two-dimensional grids show the level of excitation of the various fields, and the line-graphs under the grids indicate the intensity and localization along the retinal angle axis of excitation in the inhibitory pools. (House, 1982)

4. Visuomotor Coordination in Frog and Toad

In this section, we will return to the action-oriented view of visual perception by looking at something of the behavioral repertoire of frog and toad, and asking what type of visual system could make the appropriate information available. In other words, in the title of Lettvin, Maturana, McCulloch and Pitts [1959], we seek to determine "What the Frog's Eye tells the Frog's Brain". The key to the Lettvin, *et al*. paper was to use naturalistic stimuli such as small moving objects similar to the frog's prey of flies and worms. Indeed, extending the results of Barlow [1953], they found that there were cells in the retina that seem to serve as "bug detectors", sending information back retinotopically to the tectum, the visual midbrain. They found that the retina not only sends back a map of where the bugs might be, but also sent back other maps in spatial register, including a map of where there were large dark moving objects in the visual field. It was also known that an animal without a tectum would not engage in prey-acquisition behavior. This led to the first model of prey-acquisition which involved two modules, the retina was the prey recognition module, the tectum was the module subserving motor control for prey-acquisition.

Peter Ewert conducted a number of experiments which led to a somewhat more subtle view of the situation. Firstly, he observed that removing the pretectum, a small brain region just in front of the brain tectum, yielded a toad that would snap even at large moving objects. This immediately leads to a different functional decomposition of the brain into modules. Where before we had a module for recognizing prey, now it would seem that we have two distinct modules -- one for recognizing moving objects, and one for recognizing large moving objects. It is inhibition of prey-acquisition by the latter that yields the appearance of a module for the detection of small moving objects *per se* . Further work reviewed by Ewert [1976] replaced the use of naturalistic stimuli with precise patterns to quantify the animal's response. He found that rectangles elongated in their direction of movement (worms) were stimuli of increasing effectiveness with increasing length; but an anti-worm (a rectangle moving orthogonally to its long axis) quickly became an ineffective stimulus as it was elongated; while a square had an intermediate effect, with bigger squares being more effective stimuli until a certain critical size after which they became less effective stimuli. On the basis of this, Ewert and von Seelen came up with the model of Figure 5 in which the retina was seen as a preprocessor of visual stimuli, the pretectum was a module for anti-worm recognition supplying inhibition to the motor output system, while the tectum was seen as comprising two modules, one for worm recognition and one for motor control for prey-acquisition. This motor control module was excited by the worm recognition module and inhibited by the anti-worm recognition module.

Cervantes, Lara and Arbib [1985] carried this work further by going from the modules defined purely by an overall linear response function to modules defined by the detailed interaction of neurons. Lara, Arbib and Cromarty [1982] had built on the neuroanatomical observations of Szekely and Lazar [1976] to define the "tectal column" as a basic cluster of cells working together at one locus in the tectal layers. Cervantes, et al. then modeled the tectum as in Figure 6, as an array of such columns, interfaced with an array of pretectal "anti-worm dectector" neurons, all driven by suitable classes of retinal input, and then showed that this model could explain the spatial and temporal properties of prey-predator discrimination in terms of neural interactions.

In this last model, the process of prey acquisition is subserved by three anatomically defined modules -- the retina, the tectum, and the pretectum -- but is composed of four different functional modules, at least. The tectum is not separable into distinct anatomical subsystems for its two functions of "worm filtering" and "motor control", although we can certainly discriminate the contributions that different cells make to both these functions. Moreover, we see that the interaction between these three regions is retinotopic, and mediated by different cellular pathways, so that -- for instance -- the tectum receives three classes of retinal input as well as a class of tectal input, and these inputs are not all provided to one common input layer for the tectum, but are in fact systematically distributed in layers of termination segregated at different depths and thus making contact with different cell types within the tectum.

Now let us try to embed this set of modules in a functional account for a more complex behavior. Collett has shown that there is a sense in which the toad builds a spatial map of its world. If the toad is confronted by a fence through which it can see a worm, then it will under some circumstances advance directly to snap at the worm, but in many other cases will sidestep. When it sidesteps around the fence, the direction in which it turns after sidestepping will be correlated closely with the position of the worm, even if the experimentalist has taken care to ensure that the worm is no longer visible at this time. We thus see that the animal must make a depth map which includes the position of the worm and the position of the fence, that it must use this to control a variety of motor behaviors including sidestepping, orienting and snapping. Moreover, David Ingle has provided data that shows that snapping and orienting can be dissociated by suitable lesions, so that processes to control these behaviors must be localized in different parts of the brain.

Building on these and other observations, Arbib and House [1983] have advanced a model of the interactions in the brain that subserve this, and House [1984] has further refined the models of depth perception mechanisms that are involved. The overall structure of the model is shown in Figure 7, and we will not go into any detailed exposition of it here. What we do want to stress is that there are two distinct modules for depth perception. One provides a map of the position and depth of the barriers in front of the frog. It is our hypothesis that this depth map is created based on the principles discussed in the previous section -- namely the constraint that if the world is made up of surfaces, the depth mapping algorithm can exploit this within its structure. However, when it comes to locating the position of small moving objects in space, this surface constraint is no longer viable, and in fact lesion studies have led House [1985] to come up with a very different model for prey localization, which essentially has each side of the brain select a prey target, and then use triangulation to fix its position in space. Pathways linking the two sides of the brain and involving the control of the accommodation of the eyes provide the measure of coherence that will in most circumstances ensure that each half of the brain picks the same target. What we have established here is that the output from the two eyes is incorporated not in one low-level visual representation but in two. When we consider how intimately these two are linked with the particular problems of depth and detour behavior, there is no reason to exclude the discovery of yet other visual maps driving the computation of appropriate behavior for the animal. The one other comment we make is that Figure 7 decomposes into two quite distinct parts. The retina, the depth perception maps, and the target selector all function as layered retinotopically organized stuctures. However, the motor schemas for sidestepping, orienting, snapping, etc., are no longer to be thought of as retinotopically coded. Yes, at some point the target of those movements is encoded retinotopically, but the brain stem and the spinal cord must then translate that retinotopic coordinate specification into parameters of motor neuron firing. In some sense, the motor schema is then embedded in a neural circuit which must use patterns of firing to represent parameters in a way quite distinct from the use of location of firing to specify a parameter in a retinotopic or somatopically structured layer.

With this section, then, we have not only established that cognitive science needs a modular decomposition finer than that of Fodor's, but that the choice of modules can be constrained by the data of neuroscience. Since my examples are taken from visuomotor coordination they are specific to the domain of vision, but they do violate Fodor's argument that modules have no external feedback path whereby the output of the module can affect its input.

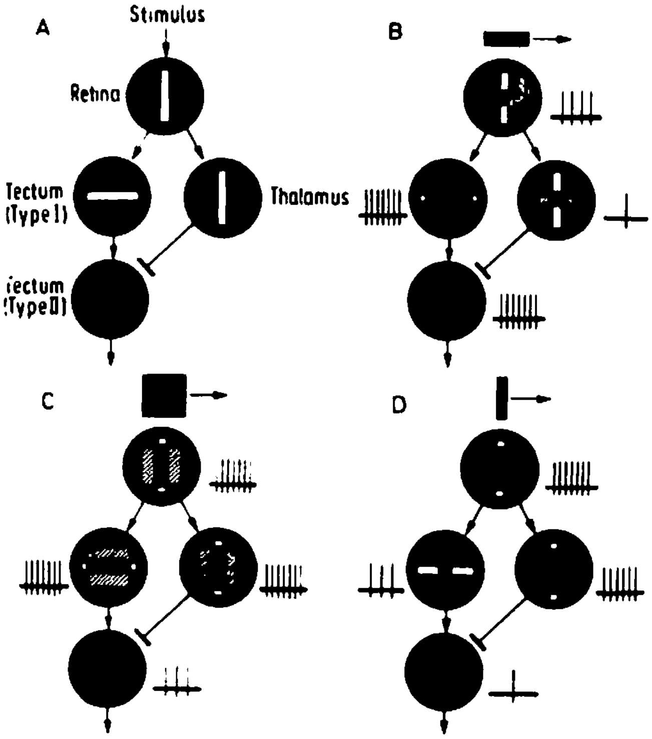


Figure 5. A lumped model of prey-predator discrimination: a worm filter provides excitatory input to an output cell which also receives inhibitory input from an antiworm cell. The result is that (B) a worm provides a potent response; (C) a square provides an intermediate response; while (D) an antiworm yields little or no response. (Ewert and von Seelen, 1974)

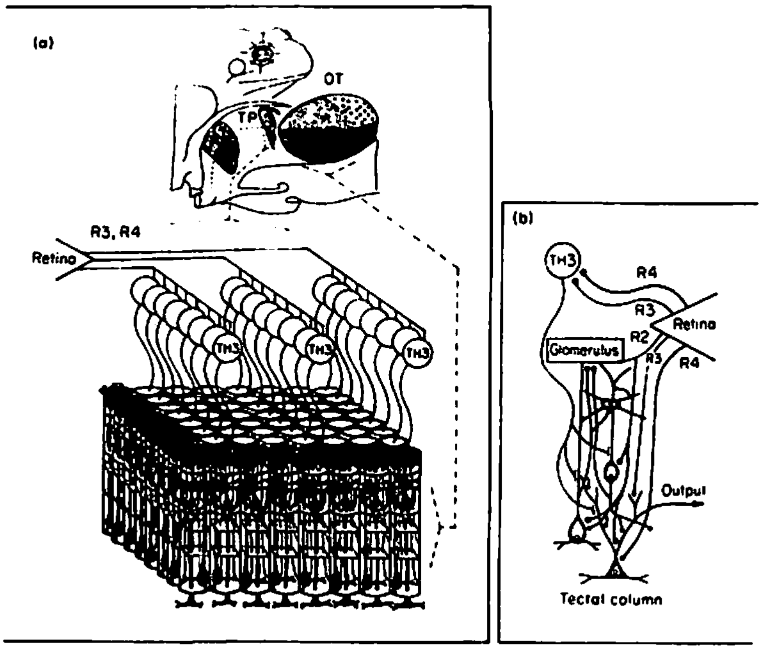


Figure 6. Interactions among retina, optic tectum and pretectum. The retina sends fibres in a retinotopical fashion to both optic tectum (class R2, R3 and R4), and pretectum (class R3 and R4). A) TH3 neurons also project retinotopically to the optic tectum. For simplicity we only show the projection of three rows of TH3 cells projecting upon the tectal columns. B) A closer look of the interactions among retinal, tectal and pretectal cells. The TH3 cell of the pretectal column inhibits LP, SP and PY of the tectal column corresponding to its retinotopic projection. (From Cervantes-Perez, Lara and Arbib (1985))

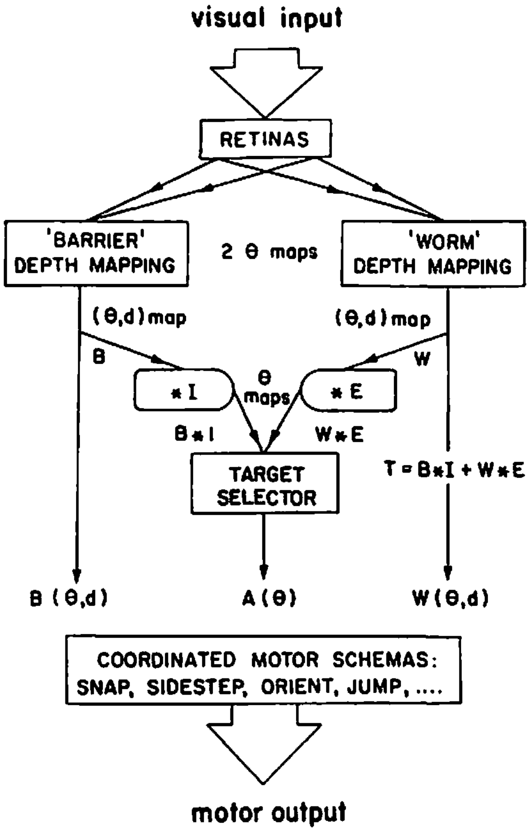


Figure 7. Conceptual Schematic of Visual/Motor Pathway: Assumptions made in this diagram are 1) that separate depth maps are maintained for prey and barrier stimuli, 2) direction for an orientation turn is obtained by combining information from these two depth maps, and 3) information on preferred orientation and depth of prey and barriers is available simultaneously to motor schemas. These schemas are capable of integrating this information to produce a coordinated motor-output. (Arbib and House, 1985)

5. The Structure of Visual Systems

In the field of machine vision, it has become commonplace to distinguish "low-level vision" from "high-level vision". Basically, low-level vision takes the retinal input and codes it into a form suitable for interpretation of, or interaction with, the world but does not depend upon knowledge of what particular objects are in the world -- though it may well depend on general properties of the world such as the hypothesis that the world is "made up of relatively few connected regions" (Section 3). By contrast, high-level vision is the process whereby this intermediate representation is used to guide the actual interpretation of the world in terms of objects, or to determine patterns of interaction with, or navigation through, the world.

In the last two sections we have given some sense of what can be done in the way of layered computation in the early stages of visual processing. In this section, I will expand our analysis of modular decomposition by showing how the module for low-level vision may be decomposed into a small number of layered submodules, while high-level vision is decomposed into the interaction of a vast number of small modules called schemas. Such schema networks embody, though I shall not argue this explicitly here, many of the Quinean and isotropic properties (see Section 6 for a definition) that Fodor holds to be a characteristic of central processes which distinguishes them from input modules.

It was J.J. Gibson who, perhaps more than anyone else, drew our attention to the immense amount of information about the structure of the world that could be inferred from the properties of surfaces and the way in which they reflected and transformed environmental energy as it passed to our receptors. However, the full impact of this work was delayed because the Gibsonians talked of "direct perception", without addressing what to most of us would seem the self-evident fact that some mechanisms within the brain must be required to carry out the inference back from those environmental energies to the distal objects which transformed them. The sort of work reviewed in section 3 marked the beginning of using the insights of Gibson but embodying it in the computational framework that he himself rejected. David Marr, as we have already mentioned, has been in the vanguard of those who would create a computational Gibsonainism, but his school has paid pehaps too little attention to the lessons that Gibson taught us. For example, Shimon Ullman, whose work on motion detection can be seen as giving computational expression to Gibsonian principles, in fact wrote a well-known article "Against Direct Perception" which perhaps let the debates against Gibson's anti-computationalism obscure the debt to him.

In this section, I want to briefly look at the issue of what representations the low-level vision can deliver, and what methods are available for interpreting it. We have already seen one example in the previous section where we discussed how low-level vision, in the form of barrier-depth mapping and prey-depth mapping, could deliver suitable information to the processes involved in detour behavior and prey acquisition. In this section, I want to turn to two approaches which represent rather the sort of processing that a human brain might go through in coming up with the recognition of the specific objects within a visual scene. I shall contrast two overall specifications, one due to David Marr, and one due to Hanson and Riseman. In the upper half of Figure 8, we see that Marr postulates a one-way flow from the primal sketch (a fancy name for the sort of feature map that we have come to see as the result of the earlier stages of visual processing, as in the bug detectors in the retina of Lettvin, et al., or the edge detectors in the cortex of Hubel and Weisel). Marr then sees the interface between low-level vision and high-level vision given by the 2-1/2 D sketch which specifies for each portion of the visual field the depth of the corresponding distal object, and the local orientation of the surface at that point. This then provides the input to the object recognition process, and Marr and Nishihara (in a suggestion which was not implemented on the computer) have suggested that many objects can be represented in terms of the connection of a variety of generalized cylinders, and have suggested how the 2-1/2 D sketch might be processed to find such cylinders, whose description could be used to key a database of known objects.

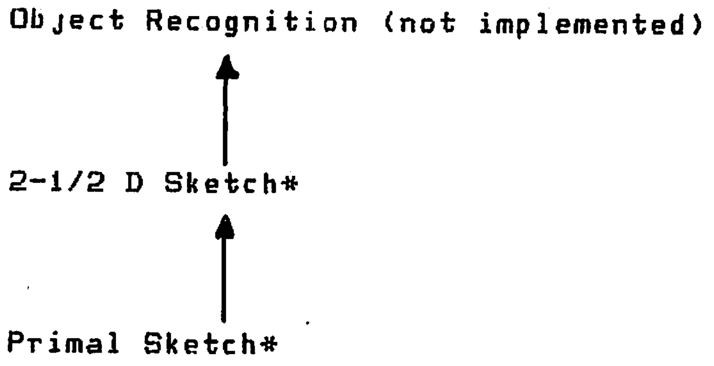
Where Marr's theory is offered as a theory of the way the human visual system must work, Hanson and Riseman's work is offered as a useful way to build a machine vision system. However, I will draw lessons from both approaches for our understanding of the human mind. First, their system has a process for extracting local features, akin to the primal sketch, but since they are working with the recognition of images provided by single color photographs, their next level is not a depth map, but is rather a segmentation of the scene into regions demarcated from each other by such cues as color and texture (lower half of Figure 8). The resultant description of the image in terms of regions of various shapes and colors and textures then provides the input for high-level vision, which through a process of cooperative excitation of various schemas leads to the final interpretation of the image, through processes which I will outline later in this section. As will I think be clear from the discussion in the previous section, it is not my intention to argue that Marr and Hanson are exclusively right in what they offer as the output map from low-level vision to high-level vision. In fact, the colleagues of each have also worked on ways of using the Gibsonian notion of optic flow to map the world in terms of its movement patterns of movement relative to the observer, and such a movement map enriches the descriptions to be offered by segmentation and depth maps. Thus, I think our correct picture is that low-level vision provides not one, but several maps, which can be used perhaps to different ends, by processes involved in image understanding or the determination of behavior. I have suggested this in Figure 9 by showing that local organizing processes and aggregation procedures can provide a symbolic representation of the segemented image without a commitment as to what that representation might be. As I have said, I think that in fact it comprises several representations, and this certainly is born out by the increasing discovery of the division of the brain into "many visual systems", as alluded to above. In the remainder of this section, I simply want to outline the approach that Hanson and Riseman have offered for high-level vision in terms of the interaction of schemas, since this approach embodies some of my own views about the sort of fine grain functional decomposition of the brain's activity that we are not yet able to follow through to detailed implementation in terms of neural networks [Arbib 1981]. I see the reconcilation of the view of functional activity in terms of schema interaction with our growing, but still limited knowledge of detailed neurophysiology as the major challenge for brain theory in the next fifteen years.

To complement the above general observations, we now present a more concrete discussion of how Hanson and Riseman orchestrate schemas in successful visual perception. Figure 10b shows the result of running segmentation algorithms on the image of Figure 10a. The top and bottom of the roof are fairly well delineated, but the left-hand edge is occluded while the right-hand edge "bleeds" into the sky and foliage. The figure also illustrates that highlighting or variation in texture may lead the algorithm to subdivide a natural region into several segments, as we see for several of the shutters. The problem, then, is to design algorithms that can, on the one hand, take a region and split it into parts that are to be seen as giving us information about different objects, and on the other hand, will aggregate regions that together characterize some distinctive portion of the image. The process of image interpretation calls on "high-level" information about possible objects in the scene. For example, information about houses would, amongst other things, initiate a search for a near-parallelogram as candidate for the roof-image. However, the program would not fail if there were no parallelogram in the image, but might pursue more subtle, possibilites, e.g. "if you find two approximately parallel lines at top and bottom and portions of approximately parallel lines on the left and right, join up the lines, and explore the hypothesis that the resultant parallelogram is a roof." Given a confident roof hypothesis, the system can hypothesize that below the roof the image will contain shutters or windows. Thus if regions there can be aggregated into a rectangle, the program can indeed follow the hypothesis that there is a rectangle.

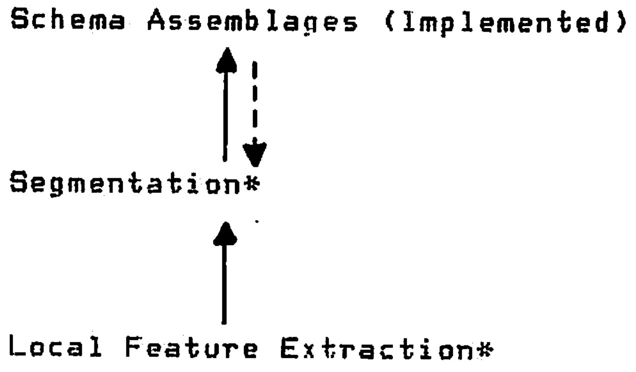
Consider Figure 11. It is not at all clear from inspection what it is, until one recognizes that it is an enlargement of the portion outlined by a small white box in Figure 10a. It is then clear that it is a bush -- we find it easy to recognize the bush in context, but hard to recognize it out of context. This suggests that, in designing a machine vision system, or in understanding the human visual system, we must understand the representation of knowledge -- the interactions between schemas -- that make this use of context possible.

Figure 12a shows a house set amongst trees, and Figure 12b shows the initial segmentation offered by the low-level vision system. A number of different schemas then work on the image to try

MARR



HANSON and RISEMAN



*Layered Computations

Figure 8.

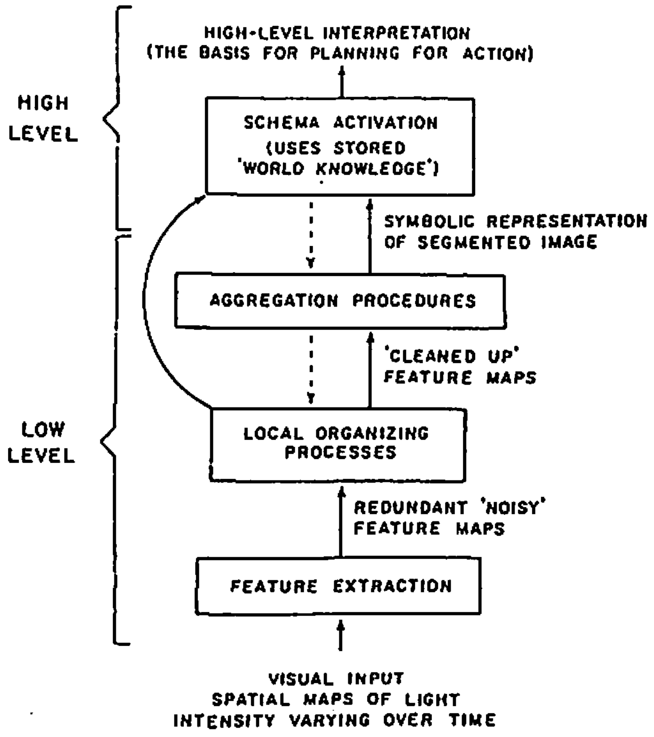

HIGH-LEVEL INTERPRETATION
(THE BASIS FOR PLANNING FOR ACTION)
HIGH LEVEL
SCHEMA ACTIVATION
(USES STORED 'WORLD KNOWLEDGE')
SYMBOLIC REPRESENTATION OF SEGMENTED IMAGE
AGGREGATION PROCEDURES
'CLEANED UP' FEATURE MAPS
LOW LEVEL
LOCAL ORGANIZING PROCESSES
REDUNDANT 'NOISY' FEATURE MAPS
FEATURE EXTRACTION
VISUAL INPUT
SPATIAL MAPS OF LIGHT
INTENSITY VARYING OVER TIME

Figure 9.

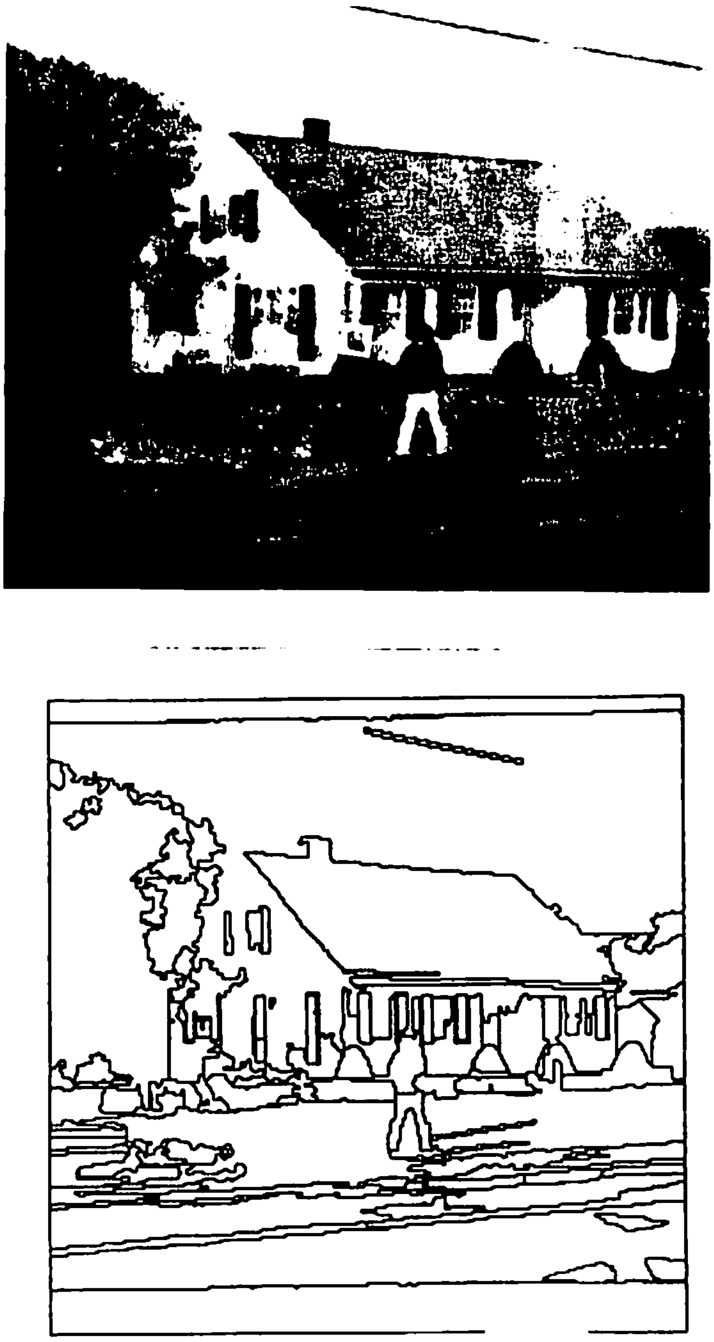


Figure 10. (a) A house image; and (b) the image segmented using low-level segmentation techniques.

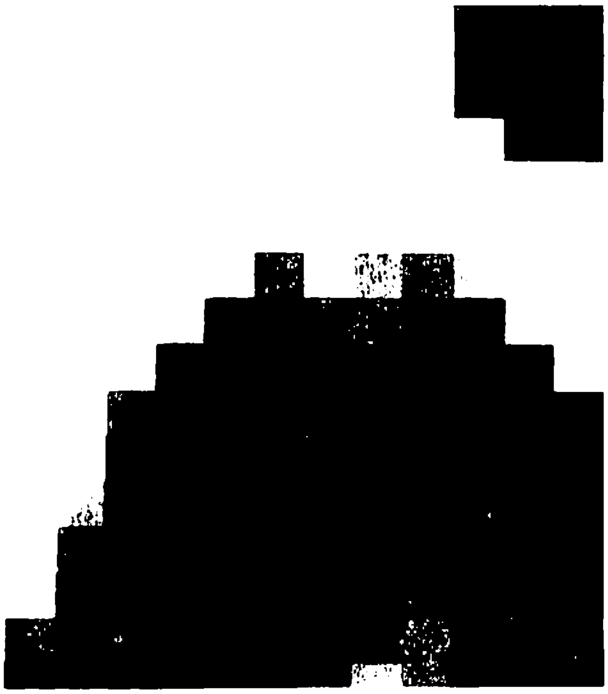


Figure 11. What is it? Compare the small box in Figure 10(a).

to find regions which can be assimilated to their corresponding objects. For example, the sky-schema will look for regions which are high in the image and which have a color in the appropriatae blue to gray range. In this case, the system comes up with a very confident assessment that what is in fact the sky-region is sky, and comes up with a lower confidence value that the roof -- which is in the upper half of the image and which has somewaht sky-like color -- could be an instance of sky. The roof schema is able to come up with a much more confident estimate that the roof region is indeed roof; it has the contextual information that the roof-region is just beneath a region which has been confidently estimated to be sky and it also has geometric information which tells it that the roof does indeed have the right shape of a partially occluded parallelogram.

At this stage we should say something about the manner of processing for the schemas. The logic is inherently parallel. Each schema can have a separate instantiation, corresponding to each region for which there is a non-trivial confidence level for the object that the schema represents. All the active schema instantiations can then communicate with each other. In the fashion that we have seen before, the activity of a schema in one region may lead to the instantiation of a schema in another region to check context. If the newly activated schema does indeed find the expected context, then the confidence level of the original schema can be raised. Although research is under way on implementing schemas on computer networks, most work is currently done by simulating the interaction of schemas on a serial computer. Thus, a number of scheduling strategies may be imposed upon the logical parallelism of the schemas. However, in what follows, we will not discuss these present implementation details, but rather stress a number of the properties of schema-based interpretation.

Returning to the present example, the high confidence level of the roof-schema activates the house-schema then searches for walls which, in both senses of the word, support the roof. The confidence level for the wall-schema can be increased for a region which is not only beneath the roof but also contains rectangular "cutouts" which could correspond to shutters, doors, or windows.

The foliage-schema is activated by finding regions which meet certain color conditions; as is the grass-schema. The final interpretation, shown in Figure 12c gives a fairly accurate interpretation of the sky, foliage, roof, wall, some of the shutters, and grass in the original scene. The system has no information about the roadway, and so we see this region left uncharacterized. A number of other regions are also missed by the system.

It may be instructive, in trying to understand the use of cooperative compuation in an expert system -- in this case for schema interpretation -- to see what sort of knowledge would have to be added to the system to make it more successful. Let us focus on two aspects of the image in Figure 12a that went unnoticed when it was analyzed by the system. The first is the mailbox -- the small white rectangle atop a black post in the right-hand side of the image, about half-way up. (The mailbox is in fact a good example of how much specific knowledge must be given to a successful interpretation system. The United States of America is one of the relatively few countries in which it is common for mailboxes to be set out near the road on a post in this fashion, thus the schema that many people have for a mailbox will not include objects of this kind.) Even if the knowledge base were augmented to include a schema for mail box, it would not have been "seen" by the system as currently constituted. This is because a small region is in danger of getting "lost". This could be corrected for by having a measure of region salience, which would score how strong the contrast is between a region -- even if a small one -- and the surrounding region. Thus the distinctive white of the region would focus processing on it to find a schema which matches, rather than allow it to be subsumed as part of the grass.

Perhaps even more interesting is that one of the uninterpreted regions corresponds to a wall of the house peeking through the trees, separated from what has been interpreted as the house. Again, this region can be lost unless a measure of salience were to focus more schema activity upon it. Let us see what extra knowledge would have to be included in the system for successful processing. First, it would have to be recognized that the color of this region matched the color of the house wall, thus causing the activation of an instantiation of the wall schema to cover that region. This

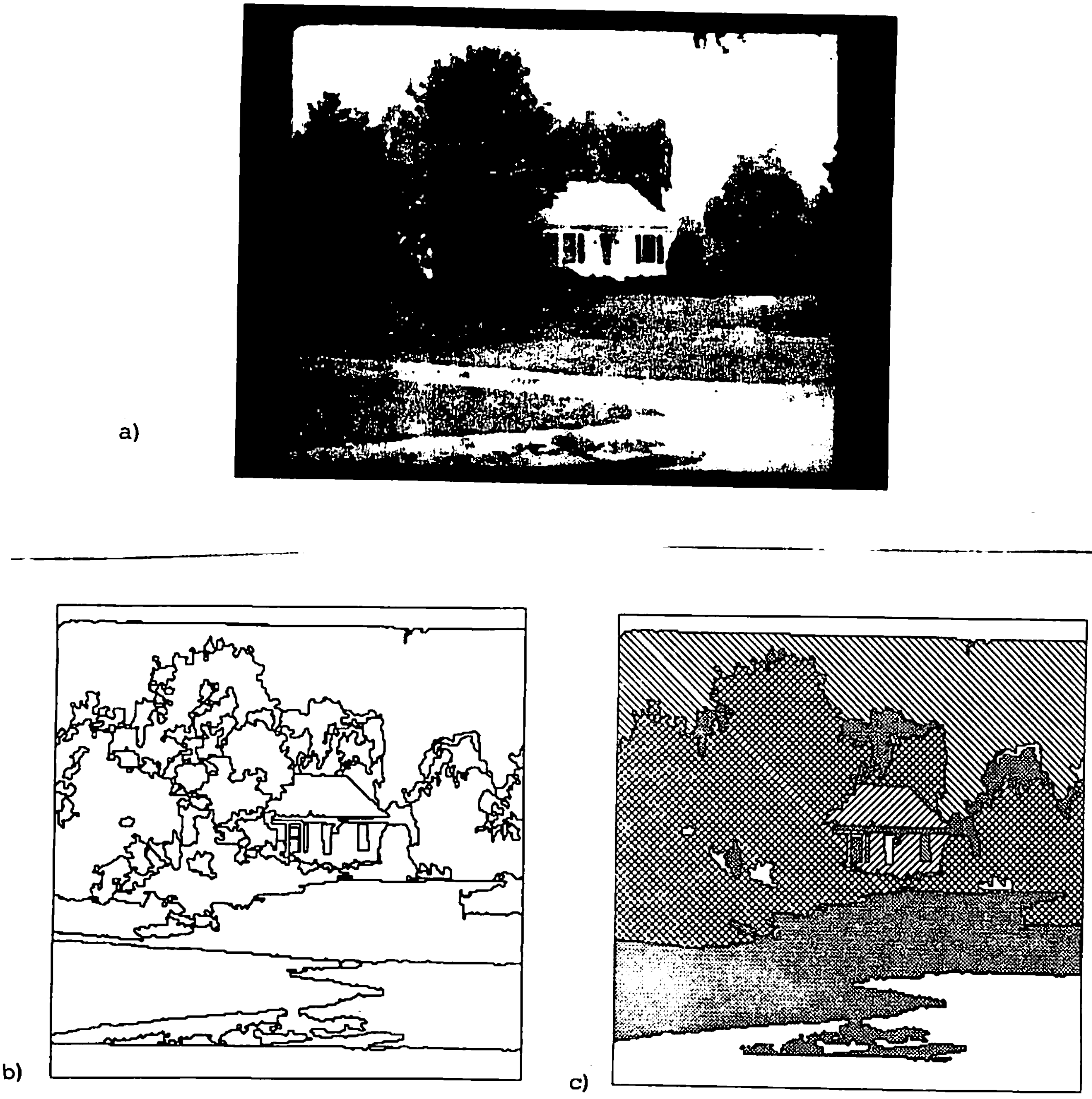


Figure 12. (a) A second house scene; (b) its segmentation; and (c) the image interpreted as a result of schema interaction.

would then cause the activation of shutter schemas to look for rectangular regions of the same color as the shutters already found in the other wall. But an interesting feature of the present image is that the color of the shutters matches so well the color of the foliage that the shutters were in fact segmented as part of the foliage. However, if a schema were looking for rectangular shape, it would be able to pull out the shutters from the foliage, to come up with a more subtle interpretation of the scene.

The second scene that we shall analyze is that we have already seen in Figure 10a, which has a segmentation (Figure 10b) which lacks a crucial edge, namely that which separates the left-hand wall from the sky. Thus the sky-schema, looking for a region which is of the right color and at the top of the image, assigns a high confidence level to the region which includes the sky and that region of the wall; while the wall-schema -- assuming that the roof-schema has already recognized the roof with high confidence -- will assign a high confidence level to regions which extend beneath the roof and which have rectangular cut-outs -- thus assigning the wall-hypothesis not only to the front wall, but to the region which embraces the side wall and the sky. Our schemas are so designed that when two different schemas are assigned a very high confidence level to the same region, they then call for low-level processes to resegment that region. It should be noted that the region of contention (Figure 13a) is very large, and that resegmenting at a greater level of detail, hopefully to find the missing edge, is a very expensive process. Here, if the schemas had more knowledge, they would be able to economize. Recognizing that the contention is between wall and sky, they would "know" that if there is a missing edge, it should be near the roofline, and would thus be able to concentrate the finer segmentation on a small region of the image -- in much the way that humans achieve with their eye movements. However, lacking this measure of "intelligence", the present implementation of the system calls for the expensive segmentation of the whole region. Subsequently, all the new sub-regions (Figure 13b), contained in what was earlier one large region will be processed by the sky-schema and the wall-schema. This time there is no problem -- those regions which are at the top of the image and of the right color become sky; while those regions which are below the roofline and abut the rectangles become interpreted as wall. The system then proceeds to the segmentation shown in Figure 13c.

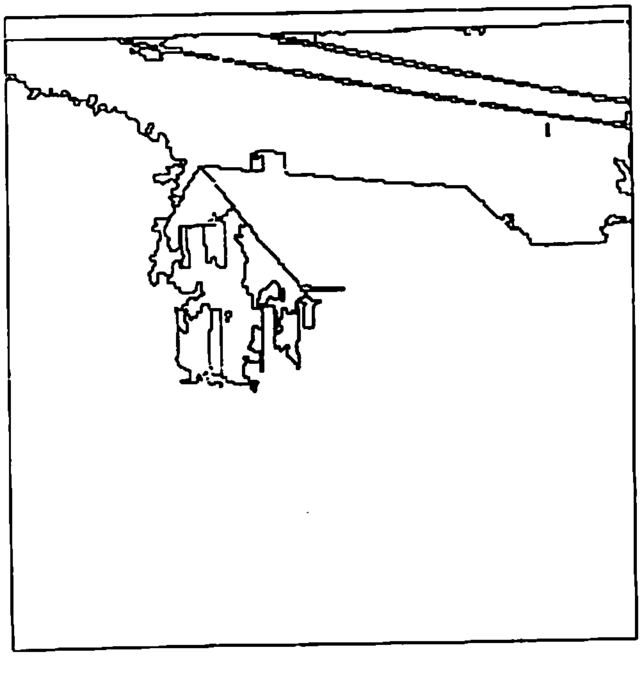
This last example makes clear two important principles;

1. Cooperative computation is not a one-way process. Although some low-level processing may be required to initiate high-level schema activity, once this schema activity is underway, it may call for the low-level processing as appropriate. In fact, in some cases, schema activity may precede low-level processing -- as when we scan the room for some object that we need.

2. Intelligence can save a lot of work. In our specific example, we saw that adding a rather small number of high-level rules would allow us to avoid a great deal of expensive, highly parallel, low-level processing.

High-level vision involves the interleaving, then, of multiple processes, a cooperative computation in which each is invoked where appropriate, possibly many times, with hypotheses being generated and discarded until the system converges on as good an interpretation as it is able to give with the facilities available to it. We claim that this style characterizes the perceptual mechanisms of brains, but it also is a useful model for "central" processes.

a)



b)

Figure 13. Because the segmentation of Figure 10(b) omitted a crucial edge, the large region of (a) is initially treated as a unit, and yields strong activation of both the sky-schema and wall schema. This conflict triggers a finer segmentation of the region, with the result shown in (b). On this basis, the system proceeds to yield the final interpretation shown in (c).

c)

6. The Prospect for Modelling Central Processes

We may agree with the notion of a module within a cognitive system as a subsystem which has a well-defined set of inputs and a well-defined set of outputs, and with a well-defined relation between inputs and outputs. This relation may involve the mediation of state variables internal to the module but the values of these state variables will only have an effect upon other modules to the extent that they determine the output of the module. However, in the preceding sections we have sought to establish the following:

a) Fodor's modules are only a subclass of modules in this sense, and his defining conditions may even be inconsistent.

b) Fodor's modules are too big, in that he would view the visual system as a single module, whereas cognitive science must offer an analysis in which the grain of the modules is at least as fine as that of the rectangles of Figure 3, 7 or 9.

c) In the light of (b), we reject Fodor's claim that there is no external feedback path whereby the output of the module can affect its input.

d) The concept of module is not a new one initiated by Fodor, but has a long history both in functional analysis of cognitive systems and in the structural analysis of brain mechanisms of cognitive functions. The give-and-take between functional and structural (neural) decompositions of cognitive systems provides an important tool for cognitive science.

In this last section, I turn to Fodor's views on central processes. I could argue that the type of schema interactions posited for high-level vision in the previous section constitutes a viable model for the analysis of central processes, but I shall restrict myself to a critique of three dichotomies offered by Fodor as a grounding for his distinction between input processes and central processes. Rather disarmingly, Fodor admits that "there is practically no direct evidence, pro or con, on the question whether central systems are modular.... When you run out of direct evidence, you might just as well try arguing from analogies, and that is what I propose to do. [F104]". He assumes that "the typical function of central systems is the fixation of belief (perceptual or otherwise) by nondemonstrative inference. Central systems look at what the input systems deliver, and they look at what is in memory, and they use this information to constrain the computation of 'best hypotheses' about what the world is like [F104]. This leads Fodor to use scientific confirmation as the analogy to structure his model of central processes. The key to his argument is that scientific confirmation is:

> _Isotropic_: facts relevant to the confirmation of scientific hypotheses may be drawn from _anywhere_ in the field of previously established truths; and
>
> _Quineian_: the degree of confirmation assigned to any given hypothesis is sensitive to properties of the entire belief system, e.g., based on a measure of

> simplicity, plausibility or conservation which is a metric over _global_ properties of belief systems [F107-8].

Having accepted the force of his analogy, Fodor concludes that "Even if the flow of data is unconstrained within a module, encapsulation implies constraints upon the access of intramodular processes to extramodular information sources, [whereas] if...isotropic and Quineian considerations are especially pressing in determining the course of the computations that central systems perform, it should follow that these systems differ in their computational character from the [modules]" [F110-11]. And on this basis Fodor claims [F112] that the following three taxonomies are co-extensive

> FUNCTIONAL TAXONOMY: input analysis versus fixation of belief.
>
> TAXONOMY BY SUBJECT MATTER: domain specific versus domain neutral.
>
> TAXONOMY BY COMPUTATIONAL CHARACTER: encapsulated versus quineian/isotropic.

He remarks that this co-extension, if it holds, is a _deep_ fact about the structure of the mind. I shall argue not only that it does not hold, but that the individual taxonomies are not themselves useful.

Input Analysis vs. Fixation of Belief. Fodor rejects Marr's primal, 2-1/2D and 3D sketches as defining the outputs of the visual processor, since then the visual input module would not provide for the recognition of objects and events. He argues [F94-95] that basic perceptual categories constitute the output of the vision module -- these are the "middle levels" in implicational hierarchies, e.g., "dog" rather than "poodle" or "thing". In other words, it follows from Fodor's own account that the visual input module can, with appropriate visual stimulation, deliver to central processes a confident report that the distal stimulus is a dog. But in this case, no further central processing is required to fix the belief "I see a dog". Though it is incontestable that there are many beliefs whose fixation cannot be achieved by input analysis alone, Fodor's own view of the vision module denies that input analysis vs. fixation of belief constitutes a clear dichotomy.

Domain-Specific vs. Domain Neutral. Well, this has been a useful taxonomy in AI, but Fodor's use of "domain" is different, since "vision" and "language" are domains for him, but "bacterial diagnosis" or "the blocks world" apparently serve only as objects of ridicule. Fodor asserts [F103; F139, Note 43] that "Early AI tried to treat central processes as though they were modular. Intellectual capacities were divided into...arbitrary subdepartments...and the attempted simulations proceeded by supplying machines with very large amounts of more or less disorganized, highly topic-specific facts and heuristics. ... What emerged was a picture of the mind that looked embarassingly like a Sears catalogue ... I take it that the bankruptcy of this sort of AI is self-evident and constitutes a strong prima facie argument that the organization of central processes is not modular".

However, it is not self-evident to me that this approach is bankrupt. I think that many cognitive scientists would agree that it is only by using a "microworld" to reduce problems of ambiguity and to limit the amount of information to be represented in a model that we can make the progress in analyzing process and representation which can provide the inductive base for the inference of more general mechanisms. However, a more telling point in the context of Fodor's argument is that scientific confirmation, his key analogue for central processes, itself "looks like a Sears catalogue". There are scientific societies for the study of physics, biology, chemistry, etc., and the volume of scientific research has grown to such an extent that the work of most scientists is confined to narrow subdisciplines with little communication between them: within cognitive science, work in linguistics is little affected by work in vision; within linguistics, a researcher in language acquisition may pay little heed to studies of historical phonology. We may decry this specialization, but it is neither "embarassing" nor "self-evidence of bankruptcy". If we accept AI usage, then the domain-specific vs. domain-neutral dichotomy does not match Fodor's other dichotomies, for an AI model of a central process will often span the dichotomy, using a domain-neutral "inference engine" to process facts from a "domain-specific" database -- and many of us would find even this dichotomy to be mistaken, arguing that knowledge and processes are often intertwined. If we insist on Fodor's usage of domain, we find it is not defined, but is given by a list "vision, language,..." whose continuation is unclear. Is "reading" a domain, or is it to be regarded as "sort-of-domain-specific", involving language and vision but not certain other domains, or is it "domain-neutral" since neurological studies show it to involve many interacting submodules, including ones linked to vision, language and hearing? In any case, the dichotomy has become a trichotomy. But the situation is even worse than this. Lacking any clear definition of domain other than "like an input module", Fodor's use of his dichotomy is vacuous, and any claim that it is a deep fact that "domain-specific vs. domain-neutral" might match "input module vs. central process" is reduced to empty tautology.

<u>Encapsulated vs. Quineian/Isotropic</u>. Now that we have confronted the issue of scientific specialization, we see that this dichotomy does not usefully describe scientific confirmation. We may regard science as organized into modules with relatively restricted flows of data between them. To a first approximation, each specialty is encapsulated. However, work within the discipline is certainly Quineian and isotropic with respect to the facts and theories of the discipline itself, and only secondarily and with much longer time constants with respect to science as a whole. Since two can play at the analogies game, we might suggest that input modules correspond to scientific specialities while central processes correspond to the integrative aspects of science. If we accept this analogy, encapsulation of the input processes is at best a first approximation; if we reject it, we suggest that modularity provides a good first approximation to the structure of the central processes. In either case, we reject the claim that the encapsulated vs. Quinean/isotropic dichotomy is coextensive with the input process vs. modular process dichotomy. Moreover, we may now

counter the profound pessimism with which Fodor concludes his volume, when he argues [F137] that "...if central processes have the sort of properties that I have ascribed to them, then they are bad candidates for study". He ascribes what progress there has been in cognitive science to the fact that we have been able to study modular input systems. But if, as I argue, input processes are less encapsulated and some central processes are more encapsulated then Fodor believes, then there is no reason to accept the input process/central process taxonomy as setting a dispiriting limit to the success of cognitive science.

Bibliography

Amari, S. and Arbib, M.A. [1977] "Competition and Cooperation in Neural Nets." In: Systems Neuroscience, (J. Metzler, ed.) Academic Press, NY pp. 119-165.

Arbib, M.A. [1972] The Metaphorical Brain: An Introcution to Cybernetics as Artificial Intelligence and Brain Theory, NY: Wiley-Interscience.

Arbib, M.A. [1981a] "Visuomotor Coordination: From Neural Nets to Schema Theory." Cognition and Brain Theory 4: 23-39.

Arbib, M.A., Boylls, C.C. and Dev, P. [1974] "Neural Models of Spatial Perception and the Control of Movement." In: Cybernetics and Bionics, (Keidel, W.D., Handler, W. and Spreng, M., eds.) Oldenbourg, Munich, pp. 216-231.

Arbib, M.A. and House, D.H. [1983] Depth and Detours: "Towards Neural Models." In: Proceedings of the Second Workshop on Visuomotor Coordination in Frog and Toad: Models and Experiments, (Lara, R. and Arbib, M.A., eds.) Department of Computer and Information Science, Technical Report 83-19, University of Massachusetts/Amherst.

Arbib, M.A. and House, D.H. [1985] "Depth and Detours: An Essay on Visually-Guided Behavior." In: Vision, Brain, and Cooperative Computation (M.A. Arbib and A.R. Hanson, eds.) Bradford Books, The MIT Press, in press.

Barlow, H. [1953] "Summation and Inhibition in the Frog's Retina." J. Physiol. (Lond.) 119: 69-88.

Cervantes, F., Lara, R. and Arbib, M.A. [1985] "A Neural Model of Interactions Subserving Prey-Predator Discrimination and Size Preference in Anurans." J. Theo. Biol. 113: 117-152.

Dev, P. [1975] "Perception of Depth Surfaces in Random-Dot Stereograms: A Neural Model." Int. J. Man-Machine Studies, 7: 511-528.

Ewert, J.-P. [1976] "The Visual System of the Toad: Behavioral and Physiological Studies on a Pattern Recognition System." In: The Amphibian Visual System: A Multidisciplinary Approach (K. Fite, ed.) New York: Academic Press, pp. 141-202.

Fodor, J. [1983] Modularity of Mind (Bradford Monograph Ser.) MIT Press.

Hanson, A. and Riseman, E. [1978] VISIONS: A Computer System for Interpreting Scenes. In: Computer Vision Systems, (A. Hanson and E. Riseman, eds.) Academic Press, NY, pp. 129-163.

House, D.H. [1982] "The Frog/Toad Depth Perception System -- A Cooperative/Competitive Model." In: Proceedings of the Workshop on Visuomotor Coordination in Frog and Toad: Models and Experiments, (M.A. Arbib, ed.) Department of Computer and Information Science, Technical Report 82-16, University of Massachusetts/Amherst.

House, D.H. [1984] "Neural Models of Depth Perception in Frogs and Toads." P.h. D. dissertation, Department of Computer and Information Science, Technical Report 84-16, University of Massachusetts/Amherst.

Hubel, D.H. and Wiesel, T.N. [1974] Sequence Regularity and Geometry of Orientation Columns in the Monkey Striate Cortex, J. Comp. Neurology 158: 267-294.

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

Lara, R., Arbib, M.A. and Cromarty, A.S. [1982] "The Role of the Tectal Column in Facilitation of Amphibian Prey-Catching Behavior: A Neural Model." J. Neurosci. 2: 521-530.

Lettvin, J.Y., Maturana, H., McCulloch, W.S. and Pitts, W.H. [1959] "What the Frog's Eye Tells the Frog's Brain." Proc. IRE 47: 1940-1951.

Marr, D. [1982] Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, W.H. Freeman and Co.

Marr, D., Poggio, T. [1976] "Cooperative Computation of Stereo Disparity." Science 194: 283-287.

Marr, D., Poggio, T. [1979] "A Computational Theory of Human Stereo Vision." Proc. R. Soc. Lo. B, 204: 301-328.

Mountcastle, V.B. [1978] "An Organizing Principle for Cerebral Function: The Unit Module and the Distributed System." In: The Mindful Brain, (Edelman, G.M. and Mountcastle, V.B., eds.) The MIT Press, Cambridge, MA.

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

Powell, T.P.S. and Mountcastle, V.B. [1959] Some Aspects of the Functional Organization of the Cortex of the Post Central Gyrus of the Monkey: A Correlation of Findings Obtained in a Single Unit Analysis with Cytoarchitecture. Bull. Johns Hopkins Hosp. 105, 133-162.

Scheibel, M.E. and Scheibel, A.B. [1958] "Structural Substrates for Integrative Patterns in the Brain Stem Reticular Core." In: Reticular Formation of the Brain (Jasper, H.H., etal., eds.) Little, Brown and Company, Boston, MA pp. 31-68.

Szekely, G. and Lazar, G. [1976] "Cellular and Synaptic Architecture of the Optic Tectum." In: Frog Neurobiology (Llinas, R. and Precht, W., eds.) Springer-Verlag, pp. 407-434.

Szentagothai, J. and Arbib, M.A. [1974] "Conceptual Models of Neural Organization." In: NRP Bulletin Vol. 12, No. 3: 310-479.