VIII. FROM UNDERSTANDING COMPUTATION TO UNDERSTANDING NEURAL CIRCUITRY:

D.C. Marr and T. Poggio

Modern neurophysiology has learned much about the operation of the individual neurons but little about the meaning of the circuits they compose. The reason for this can be attributed, at least in part, to a failure to recognize what it means to understand a complex information-processing system.

Complex systems cannot be understood as simple extrapolations of the properties of their elementary components. One does not formulate a description of thermodynamic effects using a large set of wave equations, one for each of the particles involved. One describes such effects at their own level and tries to show that, in principle, the microscopic and macroscopic descriptions are consistent with one another.

The core of the problem is that a system as complex as a nervous system or a developing embryo must be analyzed and understood at several different levels. For a system that solves an information-processing problem, we may distinguish four important levels of description. At the lowest, there is basic component and circuit analysis —how do transistors, neurons, diodes, and synapses work? The second level is that of particular mechanisms: adders, multipliers, and memories accessed by address or by content. The third level is that of the algorithm, and the top level contains the theory of the overall computation. For example, take the case of Fourier analysis. The computational theory of the Fourier transform is well understood and is expressed independently of the particular way in which it is computed. One level down, there are several algorithms for implementing a Fourier transform—the Fast Fourier Transform (FFT) (Cooley and Tukey, 1965), which is a serial algorithm, and the parallel "spatial" algorithm, which is based on the mechanisms of laser optics. All these algorithms carry out the same computation, and the choice of which one to use depends upon the particular mechanisms that are available. If one has fast digital memory, adders, and multipliers, one will use the FFT; if one has a laser and photographic plates, one will use an "optical" algorithm. In general, mechanisms are strongly determined by hardware, the nature of the computation is determined by the problem, and the algorithms are determined by the computation and the available mechanisms.

Each of these four levels of description has its place in the eventual understanding of perceptual information processing and it is important to keep them separate. Of course, there are logical and causal relationships among them, but the important point is that these levels of description are only loosely related. Too often in attempts to relate psychophysical problems to physiology there is confusion about the level at which a problem arises: Is it related mainly to biophysics (like afterimages) or primarily to information processing (like the ambiguity of the Necker cube)? More disturbingly, although the top level is the most neglected, it is also the most important. This is because the structure of the computations that underlie perception depends more upon the computational problems that have to be solved than on the particular hardware in which their solutions are implemented. There is an analog of this in physics, where a thermodynamic approach represented, at least historically, the first stage in the study of matter. A description in terms of mechanisms or elementary components usually appears later.

The main point, then, is that the topmost of the four levels, that at which the necessary structure of computation is defined, is a crucial but neglected one. Its study is separate from the study of particular algorithms, mechanisms, or hardware; and the techniques needed to pursue it are new. Marr and Poggio summarize some examples of vision theories at the different levels described and illustrate the types of prediction that can emerge from each.

Examples of Computational Theories

Orientation Behavior of the Fly

The flight behavior of houseflies requires an elaborate visual flight control system. Houseflies perceive motion relative to the environment and thereby stabilize their flight course; they locate and fly toward prominent objects; they are able to track moving targets and to chase other flies; they discriminate or prefer some specific visual patterns. Work at the Max-Planck-Institut over the last few years has provided a good understanding of part of this control system, especially the approach of Reichardt and Poggio (1976), which represents an example of a theory at the topmost, computational level. The overall computation is defined and accessible to experimentation, since it involves a complete input-output transduction, from the optical input to the behavioral motor output. The theoretical description and most of the related experiments are as yet restricted to a specific part of the orientation behavior of flies. However, the theory accounts in a quantitative way for orientation, chasing, and spontaneous pattern preference behavior, at least in either one or two degrees of dynamical freedom. Although connections with physiological and anatomical data are being established, the theory is based on behavioral data. The theory leads to the following equation, which quantitatively describes and "predicts" the angular trajectory $\dot{\psi}(t)$ of a fly fixating or tracking an object moving with angular speed $\psi(t)$:

$$\theta\ddot{\psi}(t) + k\dot{\psi}(t) + k\omega\,(t) = D[\psi(t)] - r\dot{\psi}(t) + N(t) \qquad (1)$$

where the angular error $\dot{\psi}(t)$ represents the instantaneous position of the pattern on the retina of the fly. The terms on the left-hand side represent the flight dynamics (θ is the moment of inertia of the fly, k is a rotational friction constant, $\omega(t)$ is the angular speed of the object). The right-hand side describes the instantaneous torque of the fly; the term $N(t)$ is a zero-mean random process and is independent of the visual input; $r\dot{\psi}(t)$, a velocity-dependent optomotor response, is the result of a "movement computation"; $D[\dot{\psi}]$ carries the position information, acquired from the visual input by a "position computation" and represents the "attractiveness" profile associated with the specific pattern. All these terms have been characterized quantitatively through independent experiments. For instance, $D[\psi(t)]$ can be measured as the mean torque generated by a fixed, flying fly when the image of the given object is flickered (to avoid a stabilized image) on the retina at position ψ. ($-D[\psi(t)]$ is shown in Figure 68.)

Through Equation 1 (and natural extensions of it) the theory predicts a rather complex behavior: fixation of several different patterns, various instances of tracking, and some simple cases of pattern "discrimination" (Poggio and Reichardt, 1973b; Reichardt and Poggio, 1975). Figure 72 gives two examples of behavior that is quantitatively explained by this approach. In summary, the theory outlines the basic logical organization of the visual control system of the fly. It holds that the nervous system performs two main computations on the visual input, one extracting movement information (the term $r\dot{\psi}(t)$), the other providing position information (the term $D[\psi]$), and that these two terms determine the (closed-loop) behavior described by Equation 1.

There is an approximate rule for determining the $D[\psi]$ and the r that are associated with a given pattern. The rule states that the "attractiveness" $D[\psi]$ of a pattern that can be decomposed into

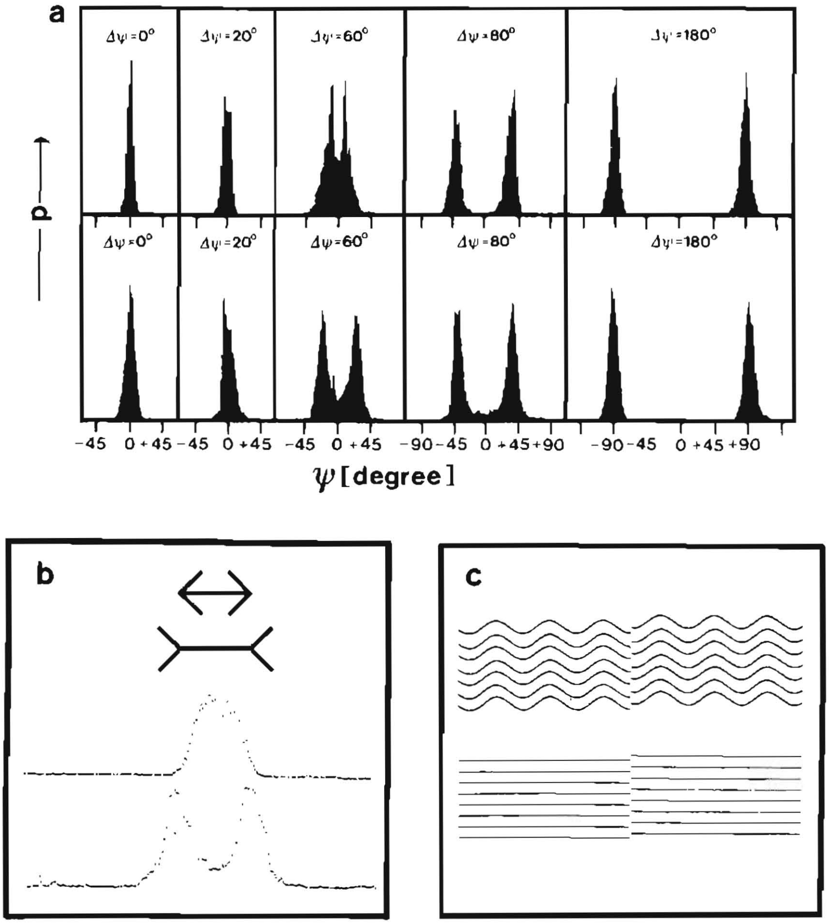


Figure 72. a. The distributions of the error angle ψ (t) during stationary fixation of two-stripe patterns (*lower figure*). The varying parameter is the angular separation of the two black, vertical stripes. *Upper figure:* corresponding histograms obtained from Equation 1. The fly fixates the center line of the pattern for separation values smaller than about 40° but either one of the two stripes for values larger than 40°. A typical phase transition occurs in the stationary fixation distribution for a value of the parameter between 40° and 60°. The nonlinear superposition of very simple, local mechanisms (see Figure 76b) leads to such a symmetry breaking. An open loop analysis (for instance via electrophysiology) cannot predict this closed loop behavior without the phenomenological theory. [Reichardt and Poggio, 1976] b. The same phase transition behavior can be observed in the fixation of the Mueller-Lyer figures. The histograms extend from −180° to 180° and show the fraction of time the fly gazed at any part of the two figures. The results conform to the theory's predictions. [Geiger and Poggio, 1975] A fly fixates (in the horizontal degree of freedom) the "illusory" vertical line arising at the boundary between the two sets of parallel horizontal lines in the upper pattern (c). There is no horizontal fixation, however, of a similar illusory, vertical line in the lower pattern. The phenomenological theory correctly predicts these results. [Poggio and Geiger, unpublished]

several vertical edges or stripe segments can be derived, to a first approximation, from the linear superposition of the "attractiveness" profiles of each component. The precise justification for this rule and its range of validity must depend on the neural interactions that compute $D[\psi]$ and $r\dot{\psi}$. A later section describes briefly this level of analysis (the "algorithm" level).

A general remark is worth making here. The quantitative description of Equation 1 could not have been obtained from single-cell recordings or from histology. This is a rather clear example where there seems to be little predictive extrapolation from the "component" level to the "computational" level. Extrapolation in the other direction is, however, somewhat easier: Equation 1 is probably a prerequisite of any full understanding at the level of circuitry. For instance, it provides a few suggestions about the modular organization of the underlying nervous network and several criteria for a physiological localization of functionally separated modules (Reichardt and Poggio, 1976).

More Complex Visual Systems

Since the early 1950's, there has been considerable progress in the study of vertebrate, and especially mammalian, visual systems. The technology of single-cell recording initiated what is widely regarded as a breakthrough in visual neurophysiology (Hubel and Wiesel, 1962). The use of the computer in psychophysics allowed Julesz (1971) to construct random-dot stereograms, and very recently it made it possible for Shepard (1975) and his collaborators to explore with precision the phenomena of "mental rotation."

Interesting and important though these findings are, one must sometimes be allowed the luxury of pausing to reflect upon the overall trends that they represent in order to take stock of the kind of knowledge that is accessible to these techniques. This *Bulletin* is itself an attempt at examining the link between the two current approaches, neurophysiology and psychophysics. We would also like to know the limitations of these approaches and how to compensate for their deficiencies.

Perhaps the most striking feature of these disciplines at present is their phenomenological character. They describe the behavior of cells or of subjects but do not explain it. What is area 17 actually doing? What are the problems in doing it that need explaining, and at what level of description should such explanations be sought?

In trying to come to grips with these problems, Marr has adopted a point of view that regards visual perception as a problem primarily in information processing. The problem commences with a large, gray-level intensity array, and it culminates in a description that depends on that array and on the purpose that the viewer brings to it. Viewed in this light, a theory of visual information processing will exhibit the four levels of description that, as we saw, are attached to any device that solves an information-processing problem; and the first task of a theory of vision is to examine the top level. What exactly is the underlying nature of the computations being performed during visual perception?

The empirical findings of the last 20 years, together with related anatomical (Zeki, 1971a,b; Allman et al., 1972, 1973; Allman and Kaas, 1974a,b,c) and clinical (e.g., Critchley, 1966; Vinken and Bruyn, 1969; Luria, 1970) experience, have strengthened a view for which widespread indirect evidence previously existed, namely that the cerebral cortex is divided into many different areas that are distinguished structurally, functionally, and by their anatomical connections. This suggests that, as a first approximation, visual information processing can be thought of as having a *modular* structure, a view that is strongly supported by evolutionary arguments. If this is true, the task of a top-level theory of vision is clear: What are the modules, what does each do, and how?

The approach of the M.I.T. Artificial Intelligence Laboratory to the vision problem rests on these assumptions. It is believed that the principal problems at present are (1) to formulate the likely modularization and (2) to understand the computational problems each module presents. Unlike the case of the fly, the first step is the most difficult, just as formulating a problem in physics is often more difficult for a skilled mathematician than solving it. Nevertheless a variety of clues is available, from psychophysics and neurophysiology to the wide and interesting range of deficits reported in the literature of clinical neurology. Those cases in which a patient lacks a particular highly circumscribed faculty are especially interesting (Efron, 1969; Warrington and Taylor, 1973); but more general impairments can also be informative, particularly the agnosias in which higher level analysis and interpretation are damaged while leaving other functions, like texture discrimination and figure-ground separation, relatively unimpaired. Such evidence must be treated with due caution, but it encourages us to examine ways of squeezing the last ounce of information from an image before taking recourse to the descending influence of high-level interpretation on early processing. Computational evidence can also be useful in suggesting that a certain module may exist. For example, Ullman (1976) showed that fluorescence may often be detected in an image using only local cues, and the method is so simple that one would expect something like it to be incorporated as a computational module in the visual system, even though there does not seem to be any supporting evidence, either clinical, physiological, or psychological. The same may be true of other visual qualities, like glitter and wetness, just as it is generally believed to be true for color, motion, and stereopsis.

In order to introduce the reader to this approach, the next few sections present brief summaries of a particular modularization and the associated theories that have been studied over the last two years. The investigators are aware that the particular decomposition chosen here may not be exactly correct, and even if it is, the separation of modules is certainly not complete. All of the modules described here have been implemented in computer programs demonstrating that this particular scheme works for a number of natural images. Alternative decompositions that have been tried, in particular those that rely on much more interaction between low-level processing and high-level interpretation of an image (e.g., Shirai, 1973; Freuder, 1975), have not hitherto led to such satisfactory and promising results.

The Primal Sketch

It is a commonplace that a scene and a drawing of a scene appear very similar, despite the completely different gray-level images to which they give rise. This suggests that the artist's local symbols correspond in some way to natural symbols that are computed out of the image during the normal course of its interpretation. The first part of the visual information-processing theory presented here therefore asserts that the first operation on an image is to *transform it into its primal sketch,* which is a primitive but rich *representation* of the intensity changes that are present, and the local geometric relations between them (Figure 74, below, shows an example). In order to obtain this description, approximations to the first- and second-directional derivatives of intensity are measured at several orientations and on several scales everywhere in the image, and these measurements are combined to form local descriptive assertions. The process of computing the primal sketch involves five important steps, the first of which can be compared with the measurements that are apparently made by simple cells in the visual cortex. One prediction made by this part of the theory is that a given intensity change itself determines which simple-cell measurements are used to describe it. This is in direct contrast to theories that assert that each simple cell acts as a "feature-detector" whose output is freely available to subsequent processes. If this is true, it requires that a well-defined interaction take place between simple-cell-like measurements made at the same orientation and position in the visual field but with different receptive field sizes (Marr, 1976c). Geometrical relations between neighboring items in the image are represented in the primal sketch by "virtual" lines joining them (Marr 1976c).

Stereopsis

Suppose that images of a scene are available, taken from two nearby points at the same horizontal level. In order to compute stereoscopic disparity, the following steps must be carried out: (1) a particular location on a surface in the scene must be chosen from one image; (2) that location must be identified in the other image; and (3) the relative positions of the two images of that location must be measured. Notice that methods based on gray-level correlation between images fail to satisfy these conditions because a gray-level measurement does not define a point on a physical surface independently of the optics of the imaging device. The matching must be based on objective markings that lie on a physical surface, and so one has to use predicates that correspond to changes in reflectance. One way of doing this is to obtain a primitive description of the intensity changes that exist in each image and then to match these descriptions. Line and edge segments, blobs, and edge termination points correspond quite closely to boundaries and reflectance changes on physical surfaces.

The stereo problem may thus be reduced to that of matching two primitive descriptions, one from each eye. One can think of elements of these descriptions as having only position information, like the black points in a random-dot stereogram, although in practice there exist some rules about which matches between descriptive elements are possible, and some which are not. There are physical constraints that translate into two rules for how the left and right descriptions are combined: (1) *The uniqueness condition.* Each item from each image may be assigned at most one disparity value. This condition rests on the premise that the items to be matched have a physical existence and can be in only one place at a time. (2) *Continuity.* Disparity varies smoothly almost everywhere. This condition is a consequence of the cohesive-

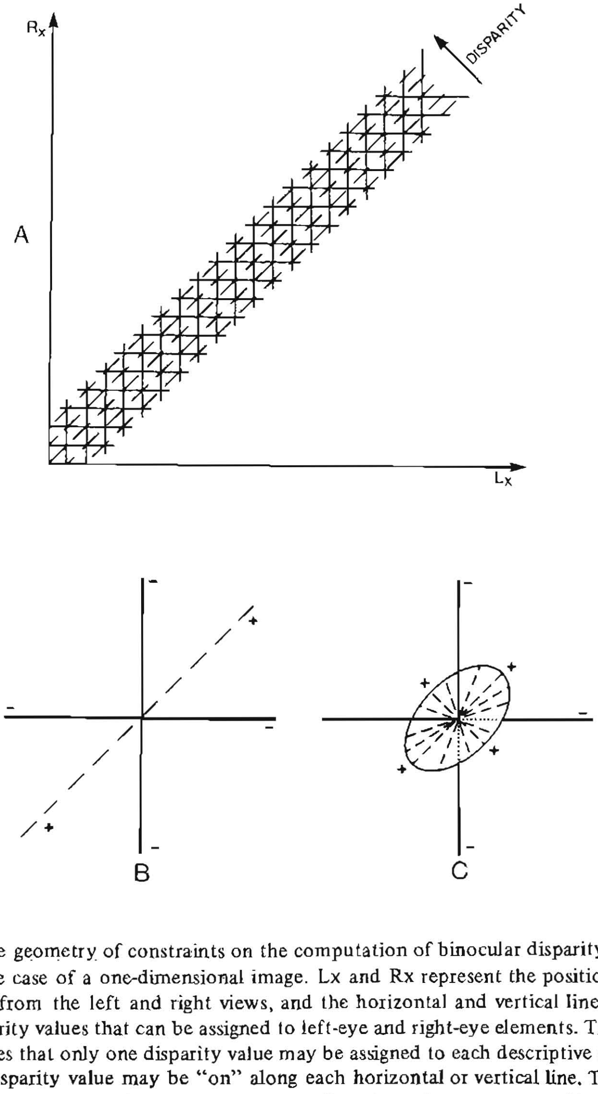


Figure 73. The geometry of constraints on the computation of binocular disparity. A. The constraints for the case of a one-dimensional image. Lx and Rx represent the positions of descriptive elements from the left and right views, and the horizontal and vertical lines indicate the range of disparity values that can be assigned to left-eye and right-eye elements. The uniqueness condition states that only one disparity value may be assigned to each descriptive element. That is, only one disparity value may be "on" along each horizontal or vertical line. The continuity condition states that we seek solutions in which disparity values vary smoothly almost everywhere. That is, solutions tend to spread along the dotted diagonals, which are lines of constant disparity, and between adjacent diagonals. B shows how this geometry appears at each intersection point. The constraints may be implemented by a network with positive and negative interactions that obey this geometry, because the stable states of such a network are precisely the states that satisfy the constraints on the computation. C. The constraint geometry for a two-dimensional image. The negative interactions remain essentially unchanged, but the positive ones now extend over a small two-dimensional neighborhood. A network with this geometry was used to perform the computation exhibited in Figure 77. [Marr and Poggio]

ness of matter, and it states that only a relatively small fraction of the area of an image is composed of boundaries. These conditions on the computation are represented geometrically in Figure 73A. Later, a network is exhibited that implements these conditions, and an illustration of how it solves random-dot stereograms is given.

In this case the computational problem is rather well defined, essentially because of Julesz's (1971) demonstration that random-dot stereograms, containing no monocular information, yield stereopsis. It is not yet completely clear, however, what mechanisms are actually available for implementing this computation (for instance, do eye movements play a critical role?). As a consequence, it is an open question whether the cooperative algorithm introduced later is used or whether simpler "serial" scanning algorithms may actually be implementing the stereopsis computation (Marr and Poggio, 1976).

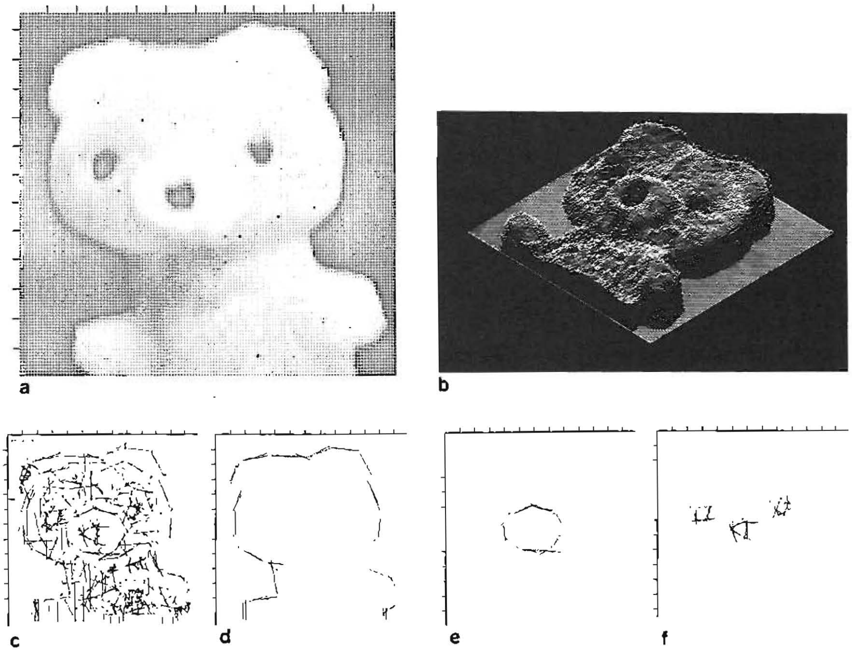


Figure 74. In a, the image of a toy bear, printed in a font with 16 gray levels, is shown. In b, the intensity at each point is represented along the z-axis. The spatial component of the raw primal sketch as obtained from this image is illustrated in c. Associated with each line segment are measures of contrast, type, and extent of the intensity change, position, and orientation. This image is so simple that purely local grouping processes suffice to extract the major forms from the primal sketch. These forms are exhibited in d, e, and f. [Marr and Poggio]

Grouping and Texture Vision

The primal sketch of an image is, in general, a large and unwieldy collection of data. This is an unavoidable consequence of the irregularity and complexity of natural images. The next important computational problem is how to decode the primal sketch. For most images, it appears unnecessary to invoke specific hypotheses about what is there until considerably later in the processing. The theory next applies a number of general selection and grouping processes to elements in the primal sketch. The purpose of these processes is to organize the local descriptive elements into forms and regions, which are closed contour groups that are obtained in various ways. Regions may be defined by their boundaries, which have been formed by grouping together some set of edge, line, or place-tokens; or they may be defined by a first-order predicate operating on the primal sketch elements within it. This second method corresponds to the definition of a region by a texture, and it leads to a theory of the processes on which texture discrimination is based.

It is important to realize that the descriptive items that may be grouped here can be very abstract—like tokens for the end of a line, a blob, or a constructed line that joins two blobs. Tokens are created for each new group, and these tokens themselves become subject to the operation of the same or similar grouping processes as operated on elements of the raw primal sketch. The grouping processes are very conservative. They satisfy a principle that seems to have general application to recognition problems, called the *principle of least commitment,* which states that nothing should be done that may later have to be undone. Only obvious groupings are made, and where there is doubt between two possible groupings, both are constructed and held pending subsequent selection. Figure 74 illustrates some results of applying these grouping processes.

3-D Representation of Shape (Marr and Nishihara, 1977)

The last two components of the theory concern the representation of three-dimensional shapes. One component deals with the nature of the representation system that is used, and the other with how to obtain it from the types of description that can be delivered from the primal sketch. The key ingredients of the representation system are:

1. The deep structure of the three-dimensional representation of an object consists of a stick figure, where in formal terms each stick

represents one or more axes in the object's generalized cone representation, as illustrated in Figure 75. In fact, a hierarchy of stick figures exists that allows one to describe an object on various scales with varying degrees of detail.

2. Each stick figure is defined by a propositional data base called a *3-D model.* The geometrical structure of a 3-D model is specified by storing the relative orientations of pairs of connecting axes. This

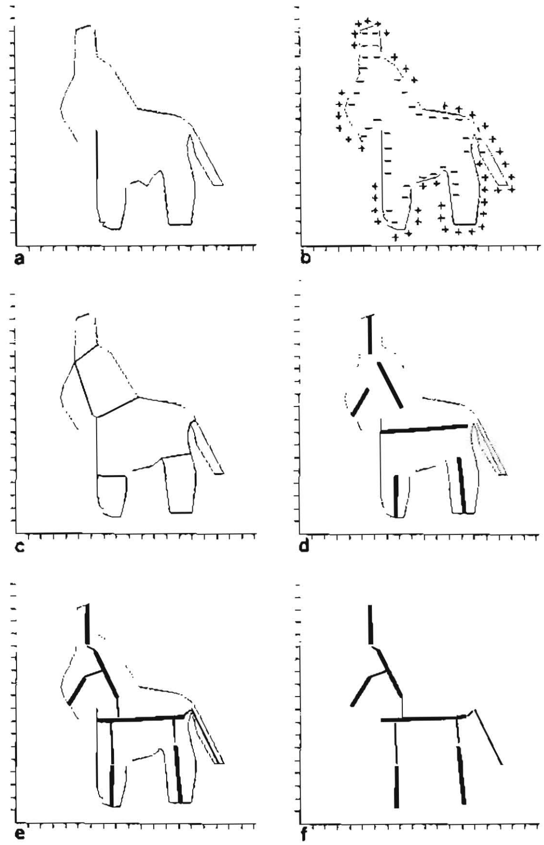


Figure 75. Analysis of a contour. The outline was obtained from a primal sketch just as Figure 74d was obtained from 74a. This contour is smoothed and then divided into convex and concave components (b). The outline is searched for deeply concave points or segments, which correspond to main segmentation points. There are usually several possible matches for each such segmentation point, but the correct mates for each may be found by eliminating relatively poor candidates. The result of this is the segmentation shown in c. Once these segments have been defined, corresponding axes are easy to obtain (d). They do not usually connect but may be related to one another by intermediate lines which are called embedding relations (e). The resulting stick figure is shown in f, which, according to the theory, is the deep structure on which interpretation of this image is based. [Marr and Poggio]

specification is local rather than global, and it contrasts with schemes in which the position of each axis is specified in isolation, using some circumscribing frame of reference.

3. When a 3-D model is being used to interpret an image, the geometrical relationships in the model are interpreted by a computationally simple mechanism called the *image-space processor,* which may be thought of as a device for representing the positions of two vectors in 3-D space, and for computing their projections onto the image.

4. During recognition, a sophisticated interaction takes place between the image, the 3-D model, and the image-space processor. This interaction gradually relaxes the stored 3-D model onto the axes computed from the image. Some facets of this process resemble the computation of a 3-D rotation, but a simple computer graphics metaphor is misleading. In fact, the rotations take place on abstract vectors (the axes) that are not even present in the original image; at any moment, only two such vectors are explicitly represented.

The essence of this part of the theory is a method for representing the spatial disposition of the parts of an object and their relation to the viewer.

2½-D Analysis of an Image

In simple images, the forms delivered from the primal sketch correspond to the contours of physical objects. Finally, therefore, we need to bridge the gap between such forms and the beginning of the 3-D analysis described in the previous section. We call this "2½-dimensional analysis," and it consists largely of assigning to contours labels that reflect aspects of their 3-D configuration before that configuration has been made explicit. The most powerful single idea here is the distinction between convex and concave edges and contour segments. One can show that these distinctions are preserved by orthogonal projections and that they can be made the basis of a segmenting technique that decomposes a figure into 2-D regions that correspond to the appropriate 3-D decomposition for a wide range of viewing angles (see Figure 75). Marr (1976a) has proved that the assumptions that are implicit in the use of the convex-concave distinction to analyze a contour are equivalent to assuming that the viewed shapes are composed of generalized cones. This gives additional support for using the stick-figure scheme based on generalized cones to represent 3-D shapes. The theory assigns many alternating figure effects like the Necker cube to the existence of alternative, self-consistent labelings computed at this stage.

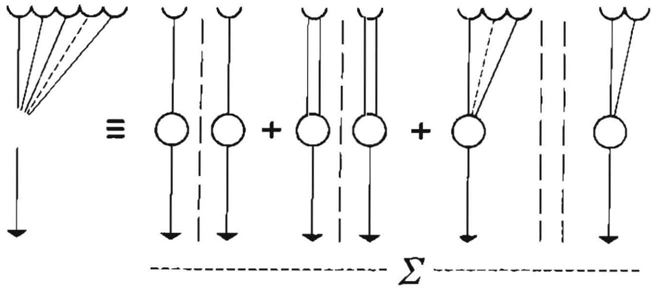
It is perhaps worth mentioning one interesting point that has emerged from this way of recognizing and representing 3-D shapes. Warrington and Taylor (1973) described patients with right parietal lesions who had difficulty in recognizing objects seen in unconventional views—like the view of a waterpail seen from above. They did not attempt to define what makes a view unconventional. According to our theory, the most troublesome views of an object will be those in which its stick-figure axes cannot easily be recovered from the image. The theory therefore predicts that unconventional views in the Warrington and Taylor sense will correspond to those views in which an important axis in the object's generalized cylinder representation is foreshortened. Such views are by no means uncommon—if a 35 mm camera is directed toward one, one sees an unconventional view of it since the axis of its lens is foreshortened.

Examples of Algorithms and Mechanisms

Between the top and bottom of our four levels lie descriptions of algorithms and descriptions of mechanisms. The distinction between these two levels is rather subtle, since they are often closely related. The form of a specific algorithm can impose strong constraints on the mechanisms, and vice versa. Let us consider three examples.

"Simple" Algorithms

An algorithm operates on some kind of input and yields a corresponding output. In formal terms, an algorithm can be thought of as a mapping between the input and the output space. Perhaps the simplest of all nonlinear operators on a linear space are the so-called polynomial operators. They encompass a broad spectrum of applications including all linear problems, and they approximate all sufficiently smooth, nonlinear operators. For this particular class of "simple" algorithms (i.e., representable through a "smooth" operator), polynomial representations provide a canonical decomposition in a series of simpler, multilinear operators (Palm and Poggio, 1977). Figure 76 shows this decomposition in terms of interactions, or "graphs," of various orders; in this way an algorithm, or its network implementation, may be dea



b

Separation of the three types of interactions in the fly

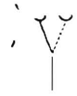
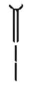
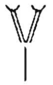
Movement computation	Position ("attractiveness") computation	
		
Corresponding to rψ	Corresponding to D(ψ)	Correction to superposition rule
Homogeneously distributed in the eye (no strong dependence on ψ and ϑ)	Mostly in the lower part of the eye (D(ψ) and L(ϑ) dependence)	Mostly in the lower part of the eye
No "age" dependence	(?)	"Age" dependence
Light intensity threshold at about 10^{-4} candel/m^2 (Eckert, 1973)	Light intensity threshold (of fixation!) at about 10^{-2} cd/m^2 (Reichardt, 1973; Wehrhahn, 1976	?
Present in the Drosophila mutant S 129 (Heisenberg, pers. comm.)	Disturbed in the Drosophila mutant S 129 (Heisenberg, pers. comm.)	?

Figure 76. Graphical representation (a) of the decomposition of a "simple" nonlinear, n-input "algorithm" into a sum of interactions of various order. The functional representation

$$S\left\{\ldots x_j(t)\ldots\right\} = L^{(0)} + \sum_i L_i^{(1)}\left\{x_i(t)\right\} + \sum_{ik} L_{ih}^{(2)}\left\{x_i(t)x_h(t)\right\} + \ldots$$

where $L^{(n)}$ is an n-linear mapping, can be read from an appropriate sequence of such elementary graphs. (b) The graphs that implement the fly's orientation behavior, studied by Reichardt and Poggio. Several findings suggest that they may correspond to separate physiological modules. Characteristic functional and computational properties can be associcted with each interaction type. [Poggio and Reichardt, 1976]

composed into an additive sequence of simple, canonical terms, just as, in another context, a function can be conveniently characterized by its various Fourier terms. Moveover, functional and computational properties can be associated with interactions of a given order and type.

Poggio and Reichardt (1976) used the polynomial representation of functions to classify the algorithms underlying movement, position, and figure-ground computation in the fly's visual system. The idea was to identify which terms, among the diversity of the possible ones, are implied by the experimental data. Figure 76 shows the graphs that play a significant role in the fly's control of flight and, in this sense, characterize the algorithms involved. The notion that seems to capture best the computational complexity of these simple, smooth mappings is the notion of p-order (perception-order, see Poggio and Reichardt, 1976). Movement computation in the fly is of order 2, and figure-ground discrimination in the simple case of relative motion depends on fourth-order graphs, but possibly with p-order 2. A closed, or Type 1 (Marr, 1976b), theory of this kind may be a useful way of characterizing preprocessing operations in nervous systems. The approach has a rather limited validity, however, since it does not apply to the large and important class of "nonsmooth" algorithms, where cooperative effects, decisions, and symbols play an essential role. While an arbitrary number of mechanisms and circuits may implement these "smooth" algorithms, it is clear that "forward" interactions between neurons are the most natural candidates.

Although the various levels of description are only loosely related, knowledge of the computation and of the algorithm may sometimes admit inferences at the lowest level of anatomy and physiology. The description of the visual system of the fly at the computational and functional level suggests, for instance, that different, separate neural structures may correspond to the various computations. Recent data support this conjecture. Movement computation (the term $r\dot{\psi}(t)$ of Equation 1 and the second order graph of Figure 76) seems to depend mainly on receptor system 1-6, while the position computation (the term $D[\psi(t)]$ of Equation 1 and the "self-graph" of Figure 76 seems dependent on receptor system 7-8 (Wehrhahn, 1976*). Mutants of *Drosophila*, normal with respect to the movement algorithm, are apparently disturbed in the position algorithm.†

"Cooperative" Algorithms

A more general and not precisely definable class of algorithms includes what one might call "cooperative algorithms." Such algorithms

*Also C. Wehrhahn, manuscript in preparation.

†M. Heisenberg, unpublished results.

may describe bifurcations and phase transitions in dynamic systems. An essential feature of a cooperative algorithm is that it operates on many "input" elements and reaches a global organization via local but highly interactive constraints. An apparently cooperative algorithm plays a major role in binocular depth perception (Julesz, 1971). The stereopsis computation defined by Figure 73A applies many local constraints to many local inputs to yield a final state consistent with these constraints. Various mechanisms could implement this type of algorithm. Parallel, recurrent, nonlinear interactions, both excitatory and inhibitory, seem to represent a natural implementation. In the stereopsis case, such a mechanism is illustrated in the rest of Figure 73. This mechanism may be realized through many different components and circuitries. In the nervous system, however, there are certain very obvious candidates, which allow some definite predictions. For instance, one is led to conjecture the existence of disparity columns (actually layers) of cells with reciprocal excitatory short-range interactions on each layer and long-range inhibitory interactions between layers with the characteristic orthogonal geometry of Figure 73. Figure 77 shows that this algorithm successfully extracts depth information from random-dot stereograms. The algorithm exhibits typical cooperative phenomena, like hysteresis and disorder-order transitions. It is important to stress that it is the computational problem that determines the structure of the excitatory and inhibitory interactions and not "hardware" considerations about neurons or synapses. The apparent success of this cooperative algorithm in tackling the stereo problem suggests that other perceptual computations may be easy to implement in similar ways. Likely candidates are "filling-in" phenomena, subjective contours, figural reinforcement, some kinds of perceptual grouping, and associative retrieval. In fact, the associative retrieval network described by Marr (1971) in connection with a theory of the hippocampal cortex implements a cooperative algorithm.

Procedural Algorithms

Still another and larger class of algorithms is represented by the specification of procedures and the construction and manipulation of explicit symbolic descriptions. For example, the 3-D representation theory described in the previous section explains how the stick figure representation of a viewed object may be obtained from an image and manipulated during recognition. The detailed specification of the algorithms involved here is carried out by defining the data structures that are created to represent the situation and by specifying procedures that

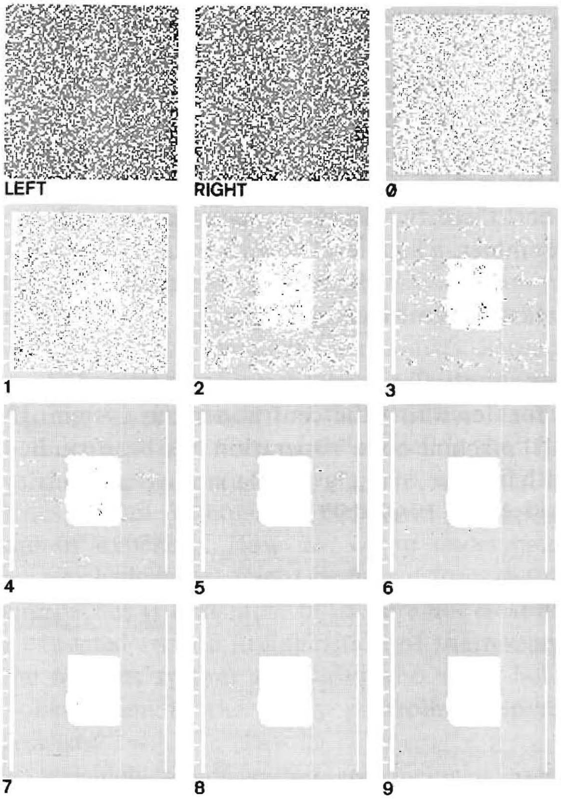


Figure 77. A pair of random-dot stereograms (left and right), the initial state of a network that implements the algorithm illustrated in Figure 73, and the first 9 iterations of the network operating on this stereo pair. To understand how the figures represent states of the network, imagine looking down on it from above. The different disparity layers in the network are in parallel planes spread out horizontally, and the viewer is looking down through them. In each plane, some nodes are on and some are off. Each layer in the network has been assigned a different gray level, so that a node that is switched on in the lowest layer contributes a dark point to the image and one that is switched on in the top layer contributes a lighter point. Initially (iteration 0) the network is disorganized, but in the final state order has been achieved (iteration 9). The central square has a convergent disparity of 2 relative to the background, and it therefore appears lighter. The density of the original random-dot stereogram was 50%, but the algorithm succeeds in extracting disparity values at densities down to less than 5%. Let C denote the state of a cell (either 0 or 1) in the 3-D array of Figure 73B at the nth iteration. Then the algorithm used here reads

$$C = \mu\left\{\sum C - \alpha \sum C + \beta \sum C\right\}$$

where $\mu(x) = 0$ if $x < 0$, and $x = 1$ otherwise; S(ijh) is a neighborhood of cell (ijh) on the same disparity layer; O(ijh) represents the neighborhood of cell (ijh) defined by the "orthogonal" directions shown in Figure 73B. Excitation between parallel disparity layers may also be present. [Marr and Poggio] operate on these data structures in accordance with the information currently being delivered from the image and that available from stored models.

This way of specifying an algorithm is very general and powerful, although, unlike the two other ways discussed, it is a far cry from the circuitry level of description at which neurophysiological experiments are carried out. In a digital computer, one does not try to bridge the gap between these two levels in one step. Instead, a basic instruction set, an assembler, a high-level language (LISP, ALGOL), and a compiler are interposed to ease the burden of passing from the description of a computation down to the specification of a particular pattern of current flow.

We may eventually expect a similar intermediate vocabulary to be developed for describing the central nervous system. Hitherto, only one nontrivial "machine-code" operation has been studied in the context of neural hardware, namely simple storage and retrieval functions (Brindley, 1969; Marr, 1969, 1971).