

ment. If this is done, the representation is slightly weakened in terms of the uniqueness criterion, but not as severely as a purely viewer-centered representation would be. Another strategy is to use a shape's visible components whenever their recognition is easy but that of the overall shape is difficult. For example, the front view of a horse usually contains an excellent view of the horse's face, which can be recognized directly and provides

another route by which the horse can be recognized. This strategy will be discussed further at the end of this section. Finally, a foreshortened axis can sometimes be found from an analysis of radial symmetry in the image. A water bucket like that shown in Figure 5-9 provides an interesting example. Its principal axis and the shape about that axis are derivable by the methods discussed above for the view shown in Figure 5-9(a) but not for that in Figure 5-9(c), where the bucket's principal axis is foreshortened.

An erroneous axis is likely to be established instead, perhaps going through the flanges that attach the handle to the rim. However, a failure to produce a recognizable description with this erroneous axis would suggest that the correct axis is not the most pronounced in the image, and an alternative could be sought. The two concentric circles (made by the top and bottom rims of the bucket) are strong clues that the principal axis passes through their centers. Furthermore, because they are concentric, these circles may be at widely separated locations along that axis. Considering that possibility could lead to the desired description of the bucket even though the identity of the closer rim remains ambiguous. A local surface depth map like the 2½-D sketch, computed by means of stereopsis, shading, or texture information, is likely to play an important role in interpreting images like these.

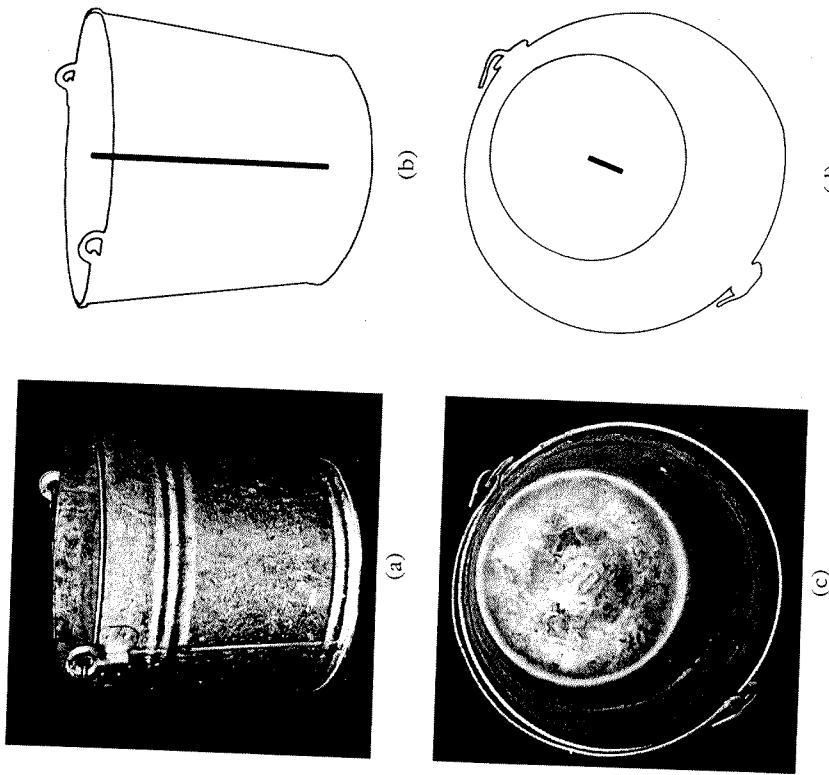


Figure 5-9. These views of a water bucket illustrate an important characteristic of any system based on the derivation of canonical axes from an image. The techniques useful for the axis shown in (b) from the image (a) are quite different from those that are best for situations where the axis is foreshortened, as in (c) and (d). (Reprinted by permission from D. Marr and H. K. Nishihara, "Representation and recognition of the spatial organization of three dimensional shapes," *Proc. R. Soc. Lond. B200*, 269-294.)

Relating Viewer-Centered to Object-Centered Coordinates

Techniques for finding axes in a two-dimensional image describe the locations of the axes in a viewer-centered coordinate system, and so a transformation is required to convert the specifications of the axes to an object-centered coordinate system. In the 3-D model representation, all axis positions are specified by adjunct relations, as in Figure 5-4, so a mechanism is required for computing an adjunct relation from the specification of two axes in a viewer-centered coordinate system. We shall call this mechanism the *image-space processor*.

The image-space processor can be kept very simple, since the adjunct relation is the only positional specification that has to be interpreted. An adjunct relation $(p, r, \Theta, \iota, \phi, s)$, as we have seen, is a way of specifying the position of a vector \mathbf{S} relative to an axis vector \mathbf{A} . What the image-space processor must do is make the coordinates of \mathbf{S} available simultaneously in a frame centered on the viewer and in one centered on vector \mathbf{A} , so that specifying the vector \mathbf{S} in either frame makes it available in the other. This is not a difficult task (see Marr and Nishihara, 1978, for more details).

The accuracy of the adjunct relations computed by the image-space processor is limited by the precision with which vectors \mathbf{A} and \mathbf{S} are specified in the viewer-centered coordinate system. Since depth information is

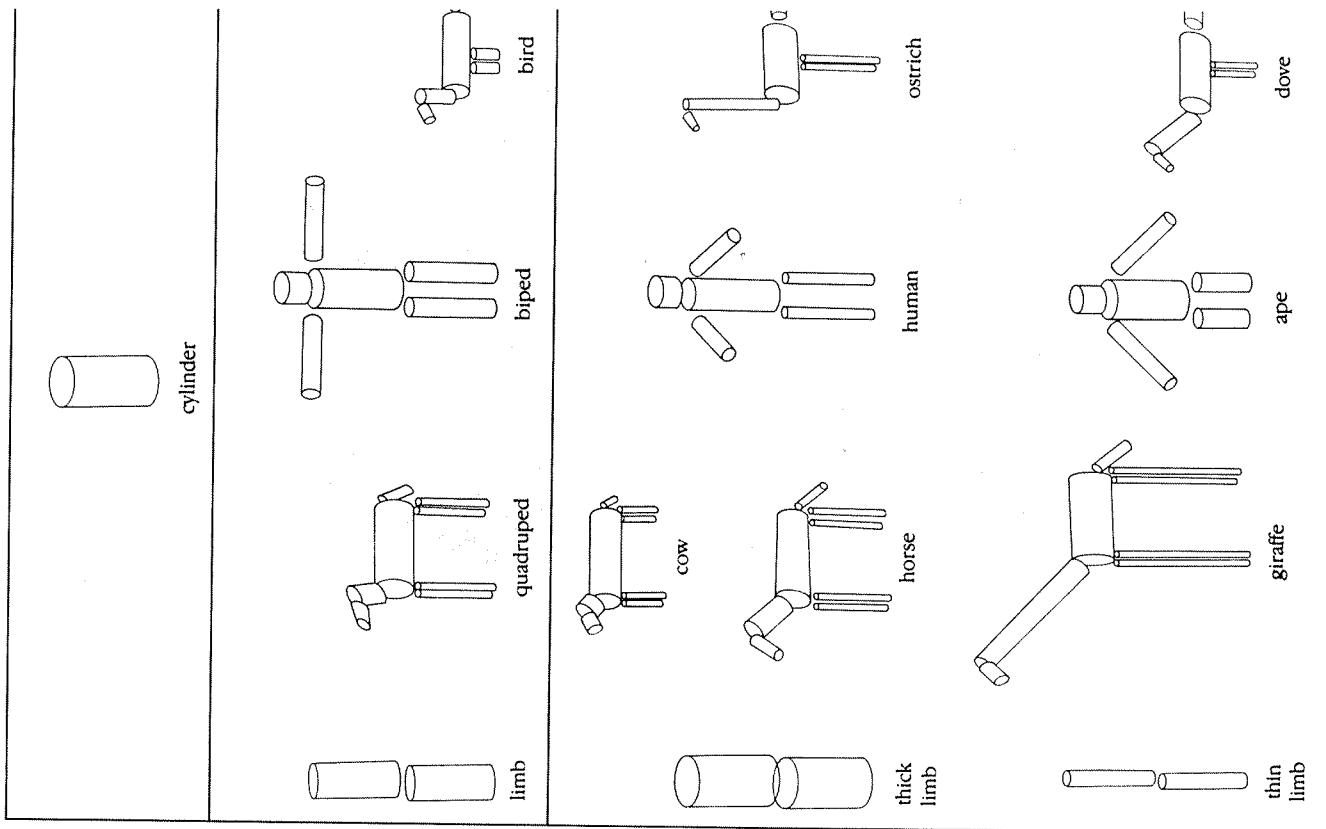
lost in the orthographic projection, the orientation specifications for axes derived from the retinal images are least precise in the amount the axes slant toward or away from the viewer. Axis slant parameters can often be reconstructed at least roughly by using stereopsis, shading, texture, structure-from-motion, and surface contour analysis. Constraints supplied by the recognition process can also be used to improve the precision of the slant specifications. We shall consider this possibility later when we discuss the interaction between the derivation process and recognition.

Indexing and the Catalogue of 3-D Models

Recognition involves two things: a collection of stored 3-D model descriptions, and various indexes into the collection that allow a newly derived description to be associated with a description in the collection. We shall refer to the above collection along with its indexing as the *catalogue of 3-D models*. Although our knowledge of what information can be extracted from an image is still limited, three access paths into the catalogue appear to be particularly useful. They are the specificity index, the adjunct index, and the parent index.

All 3-D models can be classified hierarchically according to the precision of the information they carry, and an index can be based on this classification that we call the *specificity index*. Figure 5-10 shows an exam-

Figure 5-10. (opposite) If the recognition process of relating new shape descriptions to known shapes is to be a reliable source of information about the shape, it must be conservative. This diagram illustrates an organization (or indexing) of stored shape descriptions according to their specificity. The top row contains the most general shape description, which carries information about size and overall orientation only. Since no commitment about the shape's internal structure is made, all shapes are described equally well. Descriptions in the second row include information about the number and distribution of component axes along the principal axis, making it possible to distinguish a number of shape configurations (a few are shown in this example). At this point only very general commitments are made concerning the relative sizes of the components and the angles between them. These parameters are made more precise at the third level so that distinctions can be made, for example, between the horse and cow shapes. A newly derived 3-D model would be related to a model in this catalogue by starting at the top level and working downward as far as the information in the new description allows. At that point, it could branch and form a new shape category. (Reprinted by permission from D. Marr and H. K. Nishihara, "Representation and recognition of the spatial organization of three-dimensional shapes," *Proc. R. Soc. Lond. B* 200, 260-286.)



ple of this organization for models of a few animal shapes. The top level contains the most undifferentiated description available; a 3-D model with out a component decomposition. Only the model axis is specified, so the model describes any shape. At the next level of detail are various limbs and a general quadruped shape, a biped shape, and a bird shape. These descriptions are most sensitive to the number of component axes in the model and to their distribution along the principal axis (which is the torso for most animal shapes), while only very coarse information about the lengths and orientations of the components is available. One level lower in the hierarchy the descriptions become more sensitive to angles and lengths, so that distinctions can be made between horse, giraffe, and cow shapes, for example. A newly derived 3-D model may be related to a model in the catalogue by starting at the top of the hierarchy and working down the levels through models whose shape specifications are consistent with the new model's until a level of specificity is reached that corresponds to the precision of the information in the new model.

Once a 3-D model for a shape has been selected from the catalogue, its adjunct relations provide access to 3-D models for its components based on their locations, orientations, and relative sizes. This gives us another access path to the models in the catalogue, which we call the *adjunct index*. It tells, for example, that the two similar components lying at the front end of a quadruped model are general limb models and that, for a horse model, they are more specific horse limb models. Thus the adjunct index provides useful defaults for the shapes of the components of a shape prior to the derivation of 3-D models for them from the image. It is also useful in situations where a catalogued model is not accessible via the specificity index because the description derived from the image is inadequate (perhaps because the component has very little structure).

The third access path that we consider important is the inverse of the second, and we call it the *parent index* of a 3-D model. When a component of a shape is recognized, it can provide information about what the whole shape is likely to be. For example, the catalogue's 3-D model for a horse can be indexed under each of its component 3-D models so that the 3-D model for a horse's leg provides access to the 3-D model for a horse shape.

This index would play an important role in the situation, discussed earlier, where an important axis of a shape is obscured or foreshortened. When a horse faces the viewer, the omission of the torso and hind leg axes might cause the neck axis to be selected incorrectly as the principal axis. Unless special provision has been made to handle this case, the specificity index will fail to access a horse model in the catalogue. A reasonable strategy at this point is to apply the derivation process to the components of the image. In this example, 3-D models for the head, neck, and the two

forelegs would be produced. Catalogued models for the neck and torso are likely to be found using the specificity index, and each of these would indicate via the parent index that it is a component of either the quadruped or the horse 3-D model (depending on the quality of the derived component models), providing strong evidence for considering the quadruped or horse model for the whole shape.

It is important to note that the adjunct and parent indexes play a role secondary to that of the specificity index, upon which our notion of recognition rests. We shall see below that their purpose is primarily to provide contextual constraints that support the derivation process, for example, by indicating where the principal axis is likely to be when such information cannot be obtained directly from the image. They do not prevent novel composite shapes, such as a centaur, from being described faithfully and recognized (in the case of a centaur, as a horse shape with a human bust). It may be useful to posit other indexes in the catalogue, perhaps based on color or texture characteristics (for example, the stripes of a zebra) or even on nonvisual clues, such as the sounds an animal makes, but these lie outside of the scope of this investigation.

Interaction Between Derivation and Recognition

So far, the derivation of a 3-D model has been treated separately from the process of relating that model to the stored models of the 3-D model catalogue. We view recognition as a gradual process that proceeds from the general to the specific and that overlaps with, guides, and constrains the derivation of a description from the image. After a catalogued model is selected by using one of the three indexes, we want to use it to improve the analysis of the image. There are two phases to this: First, the component axes from the image must be paired with the adjunct relations supplied by the catalogue; second, the image-space processor must be employed to combine the constraints available from the image with those provided by the model to produce a new set of derived adjunct relations that are more specific than those from the catalogue model. This last phase involves an analysis of constraints that must be satisfied by adjunct relations consistent with both the image and the information from the catalogue. The general idea of using a stored model of a shape to assist in the interpretation of an image was first used by Roberts (1965) in a computer program for producing edge descriptions of shapes built out of cubes, wedges, and hexagonal prisms from their images.

Finding the correspondence between image and catalogued model

The first phase can be thought of as a homology problem, in which the adjunct relations of a catalogue model must be related to the axes derived from an image. There may not be a complete solution. For example, the leg axes in a silhouette of a horse as viewed from the side are easily identified, but the left and right forelegs usually cannot be disambiguated without further information. Often this ambiguity may be tolerable, however, since the corresponding adjunct relations for the two legs have the same general orientation specifications (they differ only in their locations), and this is all that the following analysis makes use of.

The information available for establishing the correspondence between image and model increases as the derivation-recognition process proceeds. Initially, positional information along the principal axis of the stick figure has priority, since it is the least distorted by the perspective projection. Other clues available initially include: (1) the relative thicknesses of the shapes about the component axes (the neck of a horse is much thicker than the legs), (2) possible decompositions of component axes (the tail and legs of a horse may be roughly straight but the bust has two components that always make a large angle with one another), (3) symmetry or repetition (the legs of a horse are all the same thickness and are roughly parallel, and so have roughly the same length and orientation in the image distinguishing them from the tail), and (4) large differences in ϕ of the adjunct relation (in an image, the legs and tail of a horse usually extend to one side of the torso while the neck extends to the other). Collectively, such clues are often sufficient to relate the major components of a 3-D model to the axes derived from an image.

Homology information is also available from the adjunct and parent indexes. When a 3-D model from the catalogue is obtained by using the adjunct index, the polarity of that component's axis is automatically determined. For example, when continuing the analysis of the image of a horse to one of the legs, the polarity of the leg axis is indicated by its connection with the torso (the hoof end being distal to the junction). When the parent index is used to select a catalogue model based on the identification of some of a shape's components, the pairings for these identified components strongly constrain pairings for the remaining components. For example, in the case of a horse facing the viewer, the missing torso's location in the image can be found from the locations of the head, neck, and forelegs.

Constraint analysis

Once a homology has been established between a 3-D model and the

the possible slant angles for the axes. The basic idea is that there are often only a few combinations of the slant specifications for the projected axes in the image for which the adjunct relations derived from the image would be consistent with those supplied by the catalogue model. Or equivalently, there are often only a few orientations of the catalogue model's principal axis (relative to the viewer) for which its component axes match closely the projected axes in the image.

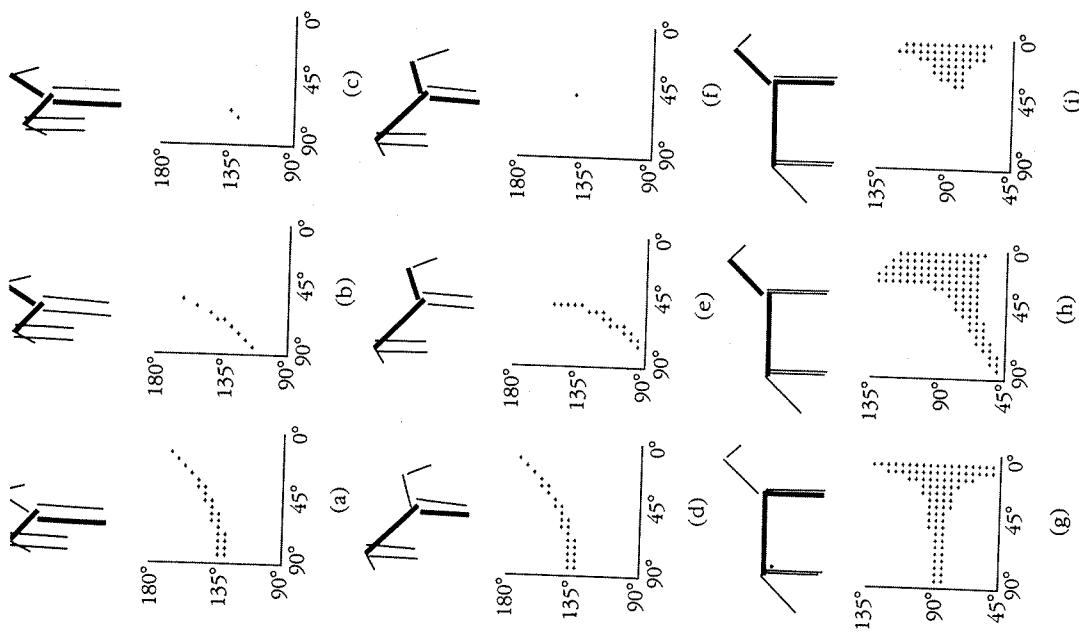
The combination of information from the image and the catalogue model is often sufficient to determine the axis slants uniquely up to a reflection about the image plane. For example, Figure 5-11(a) shows the locus of orientations of vector \mathbf{A} (relative to the viewer) that are consistent with an inclination of 90° between \mathbf{A} and vector \mathbf{S} , and an angle of 47° between their projections onto the image plane; Figure 5-11(b) shows the allowed orientations for an inclination angle of 45° and a projected angle of -111°; and Figure 5-11(c) shows the intersection of these two sets. The sharpness of these constraints depends on the particular viewing angle (as indicated by the other examples in the figure) and on the particular adjunct relations in the 3-D model. Generally, the constraints are strongest when the component axes have very different orientations and when the principal axis does not lie in the image plane.

There are several algorithms that can use these constraints. Perhaps the simplest is a relaxation process that adjusts the orientation of \mathbf{A} incrementally, seeking the disposition for which the projections of the angles between the component axes of the catalogue model, as computed by the image-space processor, best agree with those in the stick figure image. At this point vector \mathbf{A} will indicate the orientation of the principal axis that is most consistent with all of the constraints, and the image-space processor can use its other vector, \mathbf{S} , to compute the orientations of each of the component axes by using the adjuncts from the catalogue model. This hill-climbing approach converges quite efficiently when the constraints are sufficiently strong.

Alternatively, instead of relaxing the orientation of the catalogue model's principal axis, one can relax the slant angles of the sticks obtained from the image. In this case, the discrepancy measure is obtained by comparing adjunct relations derived between the sticks in the image with the corresponding adjunct relations from the catalogue. This approach is interesting because in its implementation, all of the transformations carried out by the image-space processor are in the same direction (from viewer-centered to object-centered coordinates). In a final step, improved orientation information may be used to recover more information from the image. In particular, once the orientations of the axes have been determined, their relative lengths may be computed.

The overall recognition process may be summarized as follows. We

Figure 5-11. (opposite) If we know the three-dimensional inclination angle ι that the orientation of \mathbf{A} 's coordinate system relative to the viewer is strongly constrained. (a) The orientations consistent with an inclination of 90° and an image angle like that between the heavy lines in the accompanying stick figure (allowing a tolerance of 5° in the image angle). The horizontal axis of the graph indicates the angle by which \mathbf{A} dips out of the image plane toward the viewer. The vertical axis is the amount the coordinate system is rotated about \mathbf{A} . (b) shows the set of orientations consistent with $\iota = 45^\circ$, and the visible angle between the images of the torso and neck axes. (c) The intersection of the two sets, which is restricted to a narrow range of orientations having a dip of approximately 67° out of the image plane (there is another solution, not shown here, at -67°). The remaining rows show the same analyses for dips of 45° and 0° , respectively. In this way, two-dimensional information from the image and angles from the stored 3-D model can be combined to give sometimes quite accurate information about the spatial disposition of the viewed shape relative to the viewer. (Reprinted by permission from D. Marr and H. K. Nishihara, "Representation and recognition of the spatial organization of three-dimensional shapes," *Proc. R. Soc. Lond. B* 200, 269-294.)



first select a model from the catalogue based on the distribution of components along the length of the principal axis. This model then provides relative orientation constraints that help to determine the absolute orientations (relative to the viewer) of the component axes in the image, and with this information the image-space processor can be used to compute the relative lengths of the component axes. This new information can then be used to disambiguate shapes at the next level of the specificity index.

5.6 PSYCHOLOGICAL CONSIDERATIONS

In our study of the primal sketch and of processes capable of deriving surface information from such image representations, we were much helped by evidence from neurophysiology and psychophysics and by a careful computational examination of what can in fact be derived from the available information. Our approach rested heavily on the principle of modularity (Marr, 1976), which states that any large computation should be split up into a collection of small, nearly independent, specialized sub-processes. Our analysis was based on evidence from psychophysics and from everyday experience about what the modules were likely to be, the underlying argument being that if visual information processing is not organized in a modular way, incremental changes in its design, presumably

an essential requirement for its evolutionary development, would be unable to improve one aspect of visual performance without simultaneously degrading the operation of many others. Unfortunately, we receive little help from the biological sciences about the kinds of questions raised by the later aspects of the visual process. Virtually nothing is known about the physiological and anatomical arrangements that mediate the construction of three-dimensional visual descriptions of the world, and even the best psychological information is for the most part anecdotal and derived from neurological rather than psychological studies.

Nevertheless, I think it is clear in principle that the brain must construct three-dimensional representations of objects and of the space they occupy. As Sutherland (1979) has remarked, there are at least two good reasons for this. First, in order to manipulate objects and avoid bumping into them, organisms must be able to perceive and represent the disposition of the objects' surfaces in space. This gives us a minimal requirement for something like the 2½-D sketch. Second, in order to recognize an object by its shape, allowing one then to evaluate its significance for action, some kind of three-dimensional representation must be built from the image and matched in some way to a stored three-dimensional representation with which other knowledge is already associated. As we have seen, the two processes of construction and matching cannot be rigorously separated because a natural aspect of constructing a three-dimensional representation may include the continual consultation of an increasingly specific catalogue of stored shapes.

This forces us to rely, in our study of these later problems, much more on a careful consideration of the computational and representational requirements. Stated baldly, the strong constraints come from what the representation is to be used for.

We asked here about the requirements for a shape representation to be used for recognition, and we came to three main conclusions: A shape representation for recognition should (1) use an object-centered coordinate system, (2) include volumetric primitives of various sizes, and (3) have a modular organization. A representation based on a shape's natural axes (for example, the axes identified by a stick figure) follows directly from these choices. In addition, we saw that the basic process for deriving a shape description in such a representation must involve a means for identifying the natural axes of a shape in its image and a mechanism for transforming viewer-centered axis specifications to specifications in an object-centered coordinate system.

Finally, we saw how the recognition process itself involves a mixture of straightforward derivation of shape information from the image and the

deployment of gradually more detailed stored 3-D models during the process of recognition—derivation. Thus critical ingredients of this process are a collection of stored shape descriptions and various indexes for the collection that allow a newly derived description to be associated with an appropriate stored description. The most important of these indexes allows shape recognition to proceed conservatively from the general to the specific based on the specificity of the information available from the image.

There are two ways in which we might try to examine empirically the relevance of these ideas to the processes of recognition in the human visual system. We can try to discover the type of information made explicit by the visual process in its representations, or we can try to find some correlate of the processes that derive and maintain them, perhaps using Shepard-like studies of mental rotation. The first approach is the more fundamental: Is a three-dimensional representation used, does it have a modular organization, and is it object centered? These questions have yet to be put to empirical test, but three observations are worth noting here. The first is that stick figure animals like those shown in Figure 5-1 are usually recognized easily despite the limited amount of shape information they portray. While this does not demonstrate that the human visual process is based on stick figures, it does suggest that the type of information carried by stick figures plays an important role in it.

Second, illusions like that shown in Figure 5-12 (due originally to Ernst Mach) provide evidence that local shape information is described relative to axes that are defined more globally. In the right row, the shapes are seen as diamonds, whereas along the diagonal they are seen as squares.

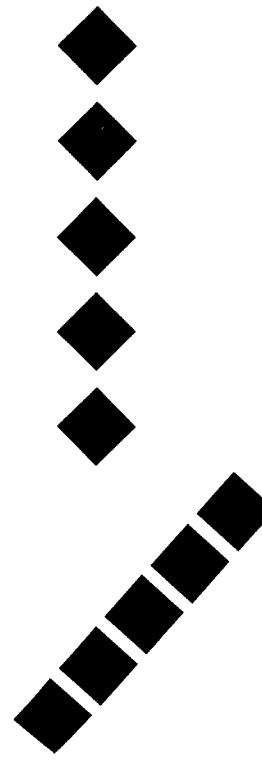


Figure 5-12. The effect of different choices of an object-centered coordinate system on the perception of shape is apparent in these diagrams. The black shapes can be seen as diamonds or squares depending on which of their several natural axes are used. (From F. Attneave, "Triangles as ambiguous figures," *Am. J. Psychol.* 81, 447–453.)

...more and more being constructed during the analysis of this pattern; it influences and therefore probably precedes the description of the shapes of the local elements.

Third, Warrington and Taylor (1973) drew attention to the difficulty experienced by their patients with right parietal lesions in interpreting certain views of common objects, which Warrington and Taylor called unconventional views. For example, these patients would fail to recognize the top view of a bucket (Figure 5-9c), denying that it was a bucket even when told it was. The patients were relatively unimpaired on views like Figure 5-9(a). As Warrington and Taylor pointed out, this difference cannot be easily explained in terms of familiarity or impaired depth perception, because both views of a bucket are common and depth is just as important to the three-dimensional structure of Figure 5-9(a) as it is to that of Figure 5-9(c). However, if the internal shape representation used for recognition was based on a shape's natural axes, the second figure would be more difficult to describe correctly, since its major axis is foreshortened. If this explanation were correct, Warrington and Taylor's unconventional views would correspond to views in which an important natural axis of the shape is foreshortened in the image, making it difficult for the patient to discover or derive a description in the shape's canonical coordinate system.

CHAPTER 6

Synopsis

Our survey of this new, computational approach to vision is now complete. Although there are many gaps in the account, I hope that it is solid enough to establish a firm point of view about the subject and to prompt the reader to begin to judge its value. In this brief chapter, I shall take a very broad view of the whole approach, inquiring into its most important general features and how they relate to one another, and trying to say something about the style of research that this approach implies. It is convenient to divide the discussion into four main points.

The first point is one that we have met throughout the account—the notion of different levels of explanation. The central tenet of the approach is that to understand what vision is and how it works, an understanding at only one level is insufficient. It is not enough to be able to describe the responses of single cells, nor is it enough to be able to predict locally the results of psychophysical experiments. Nor is it enough even to be able to write computer programs that perform approximately in the desired way. One has to do all these things at once and also be very aware of the

additional level of explanation that I have called the level of computational theory. The recognition of the existence and importance of this level is one of the most important aspects of this approach. Having recognized this, one can formulate the three levels of explanation explicitly (computational theory, algorithm, and implementation), and it then becomes clear how these different levels are related to the different types of empirical observation and theoretical analysis that can be conducted. I have laid particular stress on the level of computational theory, not because I regard it as inherently more important than the other two levels—the real power of the approach lies in the integration of all three levels of attack—but because it is a level of explanation that has not previously been recognized and acted upon. It is therefore probably one of the most difficult ideas for newcomers to the field to grasp, and for this reason alone its importance should not be understated in any introductory book, such as this is intended to be.

The second main point is that by taking an information-processing point of view, we have been able to formulate a rather clear overall framework for the process of vision. This framework is based on the idea that the critical issues in vision revolve around the nature of the representations used—that is, the particular characteristics of the world that are made explicit during vision—and the nature of the processes that recover these characteristics, create and maintain the representations, and eventually read them. By analyzing the spatial aspects of the problem of vision, we arrived at an overall framework for visual information processing that hinges on three principal representations: (1) the primal sketch, which is concerned with making explicit properties of the two-dimensional image, ranging from the amount and disposition of the intensity changes there to primitive representations of the local image geometry, and including at the more sophisticated end a hierarchical description of any higher-order structure present in the underlying reflectance distributions; (2) the 2½-D sketch, which is a viewer-centered representation of the depth and orientation of the visible surfaces and includes contours of discontinuities in these quantities; and (3) the 3-D model representation, whose important features are that its coordinate system is object centered, that it includes volumetric primitives (which make explicit the organization of the space occupied by an object and not just its visible surfaces), and that primitives of various size are included, arranged in a modular, hierarchical organization.

The third main point concerns the study of processes for recovering the various aspects of the physical characteristics of a scene from images of it. The critical act in formulating computational theories for such processes is the discovery of valid constraints on the way the world behaves

that provide sufficient additional information to allow recovery of the desired characteristic. We saw many examples of this in Chapter 3, and they were summarized in Table 3-3. The power of this type of analysis resides in the fact that the discovery of valid, sufficiently universal constraints leads to conclusions about vision that have the same permanence as conclusions in other branches of science.

Furthermore, once a computational theory for a process has been formulated, algorithms for implementing it may be designed, and their performance compared with that of the human visual processor. This allows two kinds of results. First, if performance is essentially identical, we have good evidence that the constraints of the underlying computational theory are valid and may be implicit in the human processor; second, if a process matches human performance, it is probably sufficiently powerful to form part of a general purpose vision machine.

The final point concerns the methodology or style of this type of approach, and it involves two main observations. First, the duality between representations and processes, which is set out explicitly in Figure 6-1, often provides a useful aid to thinking how best to proceed when studying a particular problem. In the study both of representations and of processes, general problems are often suggested by everyday experience or by psychophysical or even neurophysiological findings of a quite general nature. Such general observations can often lead to the formulation of a particular process or representational theory, specific examples of which can be programmed or subjected to detailed psychophysical testing. Once we have sufficient confidence in the correctness of the process or representation at this level, we can inquire about its detailed implementation, which involves the ultimate and very difficult problems of neurophysiology and neuroanatomy.

The second observation is that there is no real recipe for this type of research—even though I have sometimes suggested that there is—any more than there is a straightforward procedure for discovering things in any other branch of science. Indeed, part of the fun is that we never really know where the next key is going to come from—a piece of daily experience, the report of a neurological deficit, a theorem about three-dimensional geometry, a psychophysical finding in hyperacuity, a neurophysiological observation, or the careful analysis of a representational problem. All these kinds of information have played important roles in establishing the framework that I have described, and they will presumably continue to contribute to its advancement in an interesting and unpredictable way. I hope only that these observations may persuade some of my readers to join in the adventures we have had and to help in the long but rewarding task of unraveling the mysteries of human visual perception.

Epilogue

PART III

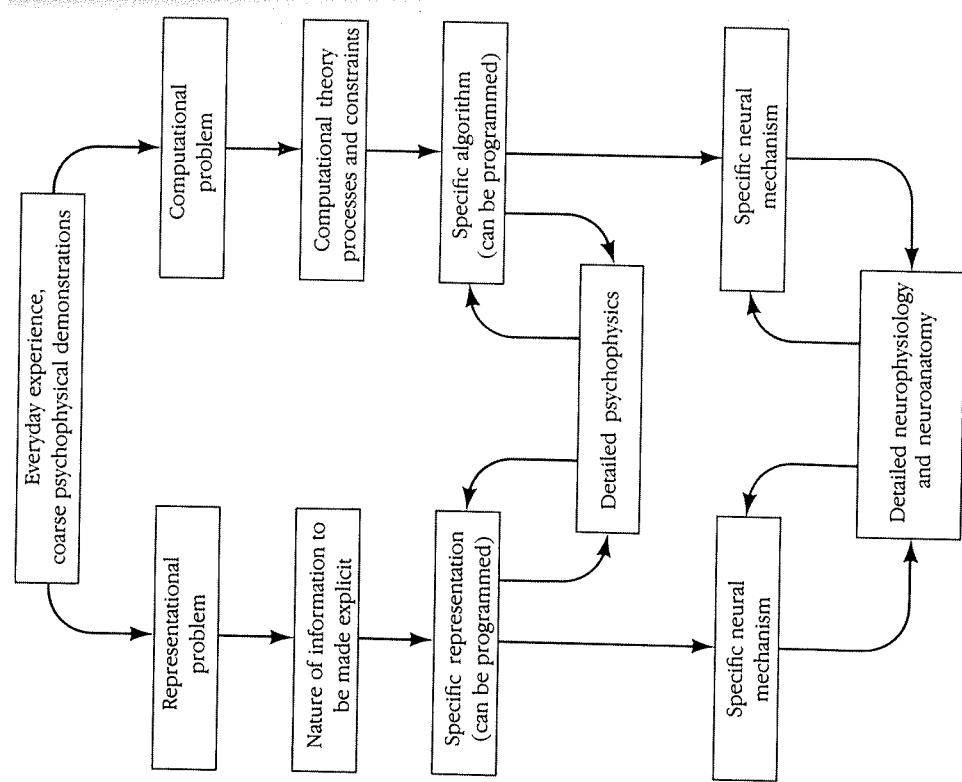


Figure 6-1. Relationships between representations and processes.

In Defense of the Approach

7.1 INTRODUCTION

In the first and second parts of this book, I have tried to set out in some detail an approach that treats visual perception primarily as an information-processing problem. I have incorporated answers to the objections most commonly raised, but from my experience in trying to convey in lectures and conversations the essence of this point of view, I expect that the reader will still have some private difficulty or question, even if it is as simple as thinking the scheme too farfetched or, at the opposite extreme, not imaginative enough.

To have addressed all the possible objections, however, would have disrupted the account too much, so I thought it best to attempt to answer them separately in the form of a conversation between an imaginary skeptic and an imaginary defender of the information-processing point of view. The dialogue is based on lunchtime conversations at the Salk Institute between Francis Crick, Tommaso Poggio, and myself, but it does not follow

those conversations very closely, and my imaginary objector is a combination of many real-life people. The discussion is not very structured and ranges over a variety of topics, but this seems unavoidable.

7.2 A CONVERSATION

Can we begin with the levels-of-explanation idea, since you attribute so much importance to it? How is it related to ideas about feature detectors and in particular to Horace Barlow's first dogma (1972, p. 380), which states, "A description of the activity of a single nerve cell which is transmitted to and influences other nerve cells, and of a nerve cell's response to such influences from other cells, is a complete enough description for functional understanding of the nervous system"?

Here, of course, I must disagree with Barlow's formulation, although I do agree with one of the thoughts behind this dogma, namely, that there is nothing else looking at what the cells are doing—they are the ultimate correlates of perception. However, the dogma fails to take level one analysis—the level of the computational theory—into account. You cannot understand stereopsis simply by thinking about neurons. You have to understand uniqueness, continuity, and the fundamental theorem of stereopsis. You cannot understand structure from motion without knowing a result like the structure-from-motion theorem, which shows how such a phenomenon is possible. In addition, and critically important for a researcher, the levels approach enforces a rigid intellectual discipline on one's endeavors. As long as you think in terms of mechanisms or neurons, you are liable to think too imprecisely, in similes.

Remember the moral from the early stereopsis networks discussed in section 3.3! None of them formulated the computational problem precisely at the top level, and almost all the proposed networks actually computed the wrong thing. Another example was the notion of segmentation to carve up an image into regions and objects. This wasted an enormous amount of time and led to the development of all kinds of special relaxation and hypothesize-and-test methods for agglomerating areas of the picture into useful regions (see Chapter 4). The problem again was that people became so entranced by the mechanisms for doing something that they erroneously thought they understood it well enough to build machinery for it—just as had occurred in the simpler case of stereopsis. It was only with a level-one attack—the formulation of the 2½-D sketch and its attendant and precisely stated problems—that real progress was possible.

Have I made my case strong enough yet? The levels idea is crucial, and perception cannot be understood without it—never by thinking just about synaptic vesicles or about neurons and axons, just as flight cannot be understood by studying only feathers. Aerodynamics provides the context in

which to properly understand feathers. Another key point is that explanations of a given phenomenon must be sought at the appropriate level. It's no use, for example, trying to understand the fast Fourier transform in terms of transistors as it runs on an IBM 370. There's just no point—it's too difficult.

For instance, take the retina. I have argued that from a computational point of view, it signals $\nabla^2 G * I$ (the X channels) and its time derivative $\frac{\partial}{\partial t} (\nabla^2 G * I)$ (the Y channels). From a computational point of view, this is a precise specification of what the retina does. Of course, it does a lot more—it transduces the light, allows for a huge dynamic range, has a fovea with interesting characteristics, can be moved around, and so forth. What you accept as a reasonable description of what the retina does depends on your point of view. I personally accept $\nabla^2 G$ as an adequate description, though I take an unashamedly information-processing point of view. A retinal physiologist would not accept this, because he would want to know exactly *how* the retina computes this term. A receptor chemist, on the other hand, would scarcely admit that these sorts of consideration have anything at all to do with the retina! Each point of view corresponds to a different level of explanation, and all must eventually be satisfied.

Yes, I see the point. You're simply saying that, from an information-processing point of view, what is done and why assumes paramount importance—this is your top level. The implementation details don't matter so much from this perspective provided that they do the right thing.

I'd like to make that point even more strongly. Figure 7-1 shows three descriptions of essentially the same thing. At the top is the mathematical piece of the retina, which we believe does roughly this, at least in part. And Figure 7-1(C) illustrates a silicon chip, built for us by Graham Nudd of the Hughes Research Laboratories in charge-coupled device technology, which carries out the $\nabla^2 G$ convolution. So, in a real sense, all these three things—the formula, the retina, and the chip—are similar at the most general level of description of their function.

Are the different levels of explanation really independent?

Not really, though the computational theory of a process is rather independent of the algorithm or implementation levels, since it is determined solely by the information-processing task to be solved. The algorithm depends heavily on the characteristics of the hardware in which it is to be implemented. For instance, biological hardware might support parallel algorithms more readily than serial ones, whereas the reverse is probably true of today's digital electronic technology.

$$\nabla^2 G * I(x,y),$$

where $\nabla^2 G(r) = -\frac{1}{\pi\sigma^4} \left(1 - \frac{r^2}{2\sigma^2}\right) \exp\left(\frac{-r^2}{2\sigma^2}\right)$

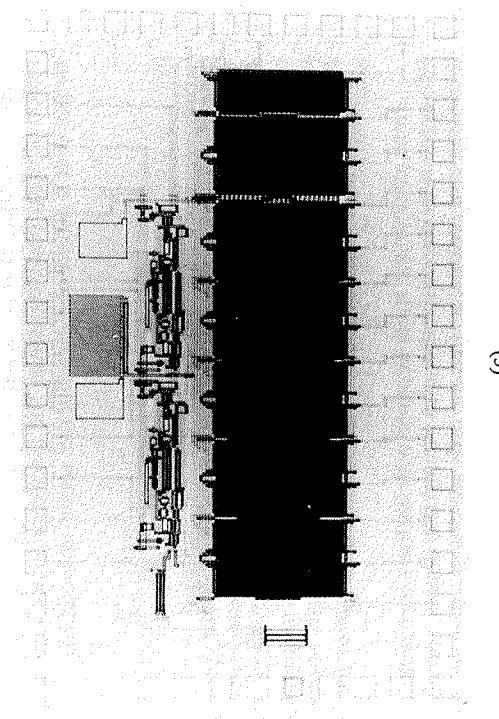
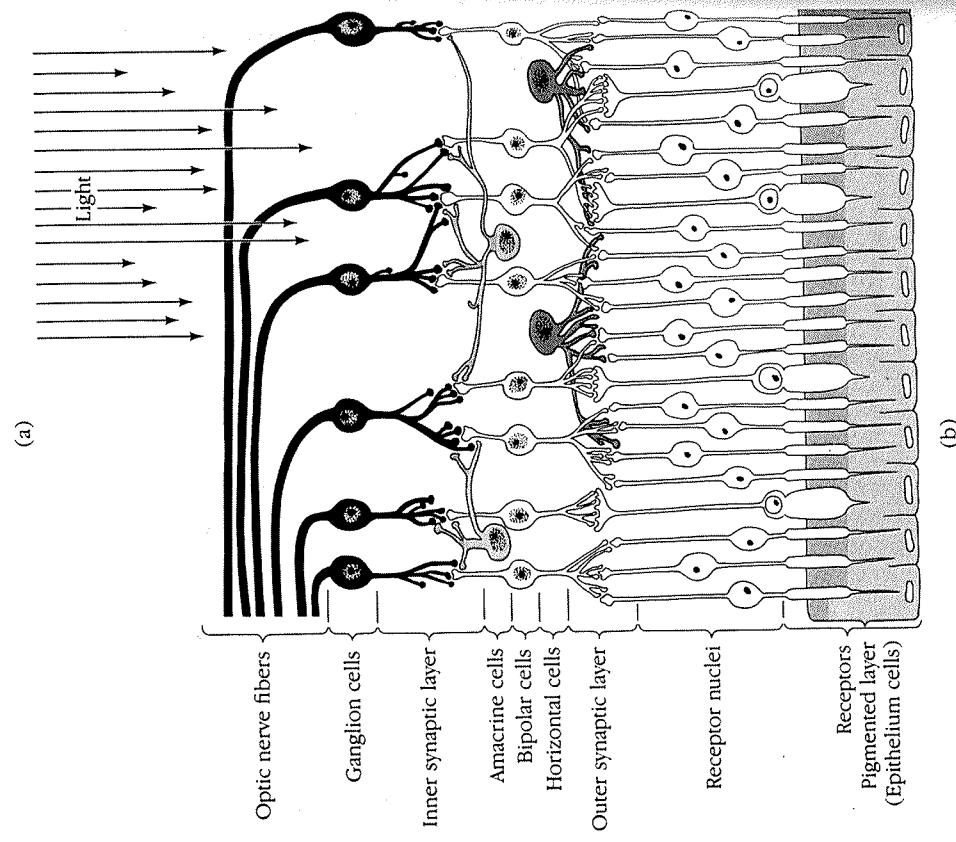


Figure 7-1 (continued).

I cannot really accept that the computational theory is so independent of the other levels. To be precise, I can imagine that two quite distinct theories of a process might be possible. Theory 1 might be vastly superior to theory 2, which may be only a poor man's version in some way, but it could happen that neural nets have no easy way of implementing theory 1 but can do theory 2 very well. Effort would thus be misplaced in an elaborate development of theory 1.

Yes, this could certainly happen, and I think it already has in the case of deriving shape from shading. I would not be at all surprised if it was unreasonably difficult to solve Horn's integral equations for shape from shading with neural networks, yet the equations can be solved on a computer for simple cases. Human ability to infer shape from shading is very limited, and it may be based on simplistic assumptions that are often violated—a sort of theory 2 of the kind you mentioned. Nevertheless, I doubt that the effort put into a deep study like Horn's was misplaced, even in the circumstances. Although it will not yield direct information about human shape-from-shading strategies, it probably provides indispensable background information for discovering the particular poor man's version that we ourselves use.

What about the old feature detector ideas? How did they fit in?
Historically, I think, the notion of a feature—and I would not now care to define it at all precisely—played an important role in shifting our conceptions away from Lashley's mass-action ideas (according to which the

brain was a kind of thinking porridge whose only critical factor was how much was working at the time) and toward the much more specific view of single-neuron action that we now have. This movement was initiated by Barlow (1953), Kuffler (1953), Lettin and others (1959), and, of course, Hubel and Wiesel (1962, 1968). Essentially, these findings ultimately lead to the notion that single nerve cells can have as one of their functions the job of signaling explicitly whenever a particular, very specific configuration is present in the input, and this type of thinking was formulated in terms of features.

But there are a number of fascinating points here arising mainly from the basic question, When does a specific configuration in the image imply a specific configuration in the environment? The first point, which we met in Chapter 1, has to do with how descriptions of the environment actually get made. In a true sense, for example, the frog does not detect *flies*—it detects small, moving, black spots of about the right size. Similarly, the housefly does not really represent the visual world about it—it merely computes a couple of parameters ($\psi, \dot{\psi}$), which it inserts into a fast torque generator and which cause it to chase its mate with sufficiently frequent success. We, on the other hand, very definitely do compute explicit properties of the real visible surfaces out there, and one interesting aspect of the evolution of visual systems is the gradual movement toward the difficult task of representing progressively more objective aspects of the visual world. The payoff is more flexibility; the price, the complexity of the analysis and hence the time and size of brain required for it.

But wasn't there more to the features idea than that?

Yes, and that, too, is an interesting set of issues that harks back to some extent to the philosophers of perception, who thought in terms of "sense atoms" grouped into larger "molecules" of sensory experience, which were the things we could recognize. One can perhaps follow a tradition of attempts at feature-based recognition. This started with the Barlow (1953) ideas, involved Kruskal's (1964) multidimensional-scaling technique, Jardine and Sibson's (1971) excellent work on cluster analysis, my early ideas about the neocortex (Marr, 1970), and the mountainous literature on statistical decision theory.

What was the main idea?

The hope was that you looked at the image, detected features on it, and used the features you found to classify and hence recognize what you were looking at. The approach is based on an assumption which essentially says that useful classes of objects define convex or nearly convex regions in some multidimensional feature space where the dimensions correspond to the individual features measured. That is, the "same" objects—members of a common class—have more similar features than objects that are not the same.

That sounds perfectly reasonable. What went wrong?

It's just not true, unfortunately, because the visual world is so complex. Does feature refer to the image or the object? Different lighting conditions produce radically different images, as do different vantage points. Even in the very restricted world of isolated, two-dimensional, hand-printed characters, it is difficult to decide what a feature should be. Think of a 5 gradually changing into a 6—a corner disappears, a gap narrows. Almost no single feature is necessary for any numeral. The visual descriptions necessary to solve this problem have to be more complex and less directly related to what we naturally think of as their representation as a string of motor strokes.

So your main argument is that the world is just too complex to yield to the types of analysis suggested by the feature detector idea?

That is correct unless, of course, the visual environment can be rigidly constrained—the lighting, the vantage point, the domain of visible elements, and so forth. If this is done, then some progress can be made. Otherwise not, and we have to look quite carefully in the literature to see this, because people do not report negative results, even though such results can be very important in deciding whether to pursue a particular line of attack.

What are the options if the domain of study cannot be so rigidly constrained?

There are basically two: Use a more complicated decision criterion or use a better representation. Using a more complicated decision criterion means abandoning the hope that classes correspond to convex clusters of features and introducing logical ideas in the decision process so that the questions asked at a given point in the classification process may depend on the answers just obtained. It is roughly true to say that artificial intelligence grew out of this approach. It leads to a view of recognition or classification as an exercise in problem solving. Decisions and routes to the solution depend sensitively on partial results found along the way, and these in turn determine the information deployed next to allow the process to continue. We saw some examples of this type of thinking in Chapter 5. The other option is to use a representation or series of representations that are better tailored to the problem at hand. In practice, this turns out to be the more important task for the particular case of vision, although for problems like medical diagnosis the problem-solving approach may be more profitable.

Are there perhaps other ways in which we might try to think about these things? What about Winograd's (1972) procedural representation of knowledge, for example, according to which terms like pick-up or block

are represented by programs. If you want to pick up the block, you simply run the two programs in sequence. That sounds like a very sensible approach to me. How does that relate to your two options?

The procedural representation idea isn't really a representation at all; it is an implementation mechanism. A representation is a much more precisely defined object. For example, there was never any result defining the scope of the procedural representation or establishing any uniqueness characteristics (in the sense of Chapter 5). It is no more a representation than is a property list! In order to define a representation, as we have seen, we must define its primitives, how they may be organized, and so on. Now the primitives in these procedural representations are simply the primitives of the underlying programming language—in Winograd's case, PLANNER or LISP. Such primitives are useless for representing what the process is actually doing in any high-level description, just as the individual instructions in a machine language program for the fast Fourier transform are useless for understanding the transform. To begin to understand and manipulate the code, one has to add comments to it. At this point it is these, not the code, that in effect provide the representation of what the code is doing from the point of view of the manipulator. G. J. Sussman's (1975) program HACKER was essentially an exercise in writing useful standard comments within a particular and restricted programming domain.

Why do you say a property list is not a way of representing knowledge? Surely it is?

I did not say that, I said it wasn't a *representation*. A property list is a programming mechanism that one may use to *implement* a representation, but it is not a representation in itself. To see this, just ask the simple question, what can and what cannot be represented in a property list, or, expressed in our earlier language, what is its scope? Is each description unique? It is meaningless to ask these questions about property lists, just as it is about procedures. Both these ideas are universal from a representational point of view, because both are in fact notions at a lower level of explanation pertaining to decisions about implementation. They are *mechanisms*, not representations. Choosing one mechanism rather than another will affect how easy it is for the programmer to make a certain piece of information explicit, but the decision about what is to be made explicit and what is not is a decision about the representation itself and is independent of the implementing mechanism.

Ab yes, and here we come back to the feature idea again. For it was surely the notion of a feature which led eventually to the idea that a representation has as its business the making of certain information explicit, wasn't it?

Very much so. But I do think that the time has now come to abandon those older ways of thinking, it being more fruitful to think instead of

systems of representations that can describe as fully as desired firstly images and then other derived aspects of the visual world. And I also think it is important not to be too anxious to relate our ideas immediately to neurons. We should first be sure that our representations and algorithms are sensible, robust, and supported by psychophysical evidence. Then we can delve into the neurophysiology.

Before leaving this topic, I feel there is one other matter we should raise. This is the question of features—well, let's call them descriptions from now on—and of measurements for getting them. What exactly is the difference between a descriptive element—perhaps we could call it an assertion—and a measurement? Is this even an important point?

There are two aspects to this. One is historical—a point I felt lay in terrible confusion back in 1974—and the second is a modern question. Let us look first at the historical question. Put most simply, people confused measurements and assertions. For example, a cell with a center-surround receptive field will respond to a blob, but it will also respond to many other things—a line, an edge, two blobs, and so forth. In fact, one can often say no more than that it signals a convolution—our old friend $\nabla^2 G * I$, for instance. Nevertheless, people did call these cells blob detectors.

Now that is not so bad in the retina, but if we were to take Hubel and Wiesel's (1962) definition of a simple cell—the simplest type of receptive field—literally, it, too, would be performing a linear convolution with one excitatory and one inhibitory strip, signalling something like a first directional derivative. I do not now believe these cells are linear convolvers (see Chapter 2), but the point is that people thought of them simultaneously as linear convolvers and as feature detectors, and that is criminal, intellectually. Of course, you can use the output of such convolvers to find edges, but it needs extra work. You have to find peaks in the first derivatives or zero-crossings in the second. And, of course, we now think that simple cells are in fact zero-crossing detectors. But the point is that here again, just because of imprecise thinking by computer vision people as well as by physiologists, that whole rich theory of early vision had been missed (see Chapter 2).

The second aspect is the modern one, and I have already raised it in Chapter 2. It has to do with when and how vision "goes symbolic." Most would agree that an intensity array $I(x,y)$ or even its convolution $\nabla^2 G * I$ is not a very symbolic object. It is a continuous two-dimensional array with few points of manifest interest. Yet by the time we talk about people or cars or fields or trees, we are clearly being very symbolic, and I think again that most would find suggestions of symbols in Hubel and Wiesel's (1962) recordings. Our view is that vision goes symbolic almost immediately, right at the level of zero-crossings, and the beauty of this is that the transition from the analogue arraylike representation to the discrete, oriented, sloped zero-crossing segments is probably accomplished without loss of information (Marr, Poggio, and Ullman, 1979; Nishihara, 1981).

And the use of symbols does not stop there either. Almost the whole of early vision appears to be highly symbolic in character. Terminations, discontinuities, place tokens, virtual lines, groups, boundaries—all these things are very abstract constructions, and few of their neurophysiological correlates have been found, but experiments like Stevens' (1978) tell us that such things must be there (see Chapter 2).

How else might one approach these phenomena? What about some kind of transformational or grammatical approach, like the one Chomsky used?

People have tried to write picture grammars involving rules that must be obeyed by line drawings (Narasimhan, 1970), but they have been unsuccessful in general and never successful on a real image. The best of the early approaches, was, I think, the blocks-world analysis of Guzman (1968), Mackworth (1973), and Waltz (1975). Unfortunately, this did not generalize—it suffered from the wrong choice of a mineworld, as indeed has much research in artificial intelligence. The great virtue of artificial intelligence has been that it forced people to substantiate their opinions by writing programs, and in doing so, these opinions were often found to be wrong. It forced a constructive way of thinking—disallowing, for example, Bertrand Russell's definition of the percept of an object as the set of all possible appearances of the object (Russell, 1921). But in having to program things, research was too often limited to a mineworld in which very many factors appear in only simple forms. Though the programs solved none of the individual problems, on the whole they ran just well enough to get by with luck. Winograd's (1972) blocks-world program was of this genre. The underlying conceptual fault is to ignore the modularity that must be present to help decompose the problem.

I do not follow. Why must it be there? How was it being ignored?

Once again, I think the clearest examples come from vision. An early mineworld, or domain of study if you like, was the blocks world—compositions of matte white prisms against a black background. The study of such a domain led to Waltz's (1975) careful cataloguing of the legal junctions of the various types of edges (as in Figure 1–3). Allowing for shadows, Waltz found that most line drawings of such scenes could be interpreted unambiguously. But notice that not one of the general processes listed in Chapter 3 was elucidated by this approach. The reason is that the general processes that combine to make up human vision cannot be easily studied by restricting oneself to any particular mineworld except by carefully choosing it in relation to something that one already suspects of corresponding to a genuine module, like the world of random-dot stereograms.

It is critical to appreciate the difference between these two kinds of mineworlds. One is very particular, the other general. Only the second kind has been found to be of value so far, although constraints in the spirit of

Waltz's may turn out to be useful for the 2½-D sketch (see Chapter 4). The reason is that for genuine computational modules with general and not limited abilities, we can actually prove theorems that show the modules will always work in the real world.

This is the true difference between the approach described in this book and the original conception of artificial intelligence, which, in its desperate effort to pack a whole working mineworld into a program—an endeavor that requires a huge amount of work—was forced to neglect and eventually to abandon attempts at real theory, turning instead to the development of better computer tools. This endeavor has met with little success. So although the artificial intelligence approach was necessary to haul us out of our false preconceptions about the simplicity of vision, it in turn became limited and hidebound because of its failure to recognize what a true computational theory is and how it should be deployed.

Are there any rules for doing this successfully?

I don't think so, and it's perfectly natural to get it wrong first. The example of flight that came up earlier makes a number of points in a nice way. First, it's obvious that you cannot understand how a bird flies by speculating on the fine structure of a feather. So the next natural step is to try to copy how the bird behaves—what I call the mimicry phase. So people built imitation wings and flapped them. That didn't work either. This phase is essentially copying at the lower two levels or possibly only at level two. The real advance comes only when you understand that an airfoil provides lift in accordance with Bernoulli's equation. That is the level-one part— aerodynamics. It is why a bird and a 747 are similar—and why both are dissimilar from a gnat, which keeps itself aloft not by means of an airfoil but by "treading air" in an essentially turbulent regime.

But at some stage, one has to relate one's level-one ideas directly to neural machinery, surely? You talked about the eyes—the retina and $\nabla^2 G$ —but what about eye movements? I understand that from your—I should say—from an information-processing and levels point of view, they are quite trivial to deal with. But that doesn't make it any easier for me to think of compensating for them in neural machinery.

Yes, I admit that this is a thorny issue. But first, I hope I made it clear in Chapter 4 that eye movements involve much more than just a subtraction. We saw there how the representation of surface orientation, for example, is quite intimately bound up with whether you choose a retinocentric polar frame (the natural one from the point of view of imaging) or a more invariant type of retinocentric frame.

The second point is that, by delaying the transition out of a retinocentric frame, the difficulty of the arithmetic that is necessary when one at last performs the transition is correspondingly eased. In the manner of Chapter

5, we can move directly to a 3-D model representation, which is located in a stable frame around the viewer; and then all we have to check is that when the eyes move, the appropriate blob moves as expected.

Lastly, I think that here, as always, it is important not to be fooled by the apparent detail and luxury of our perception. We met this earlier in connection with the immediacy and vividness of our perception. I would be surprised if we can keep track of more than a handful of objects during eye movements, and I expect our powers are quite limited in this respect.

Yes, I see the plausibility of the argument. But this doesn't need our levels, does it? It seems a rather different kind of issue.

Absolutely true, but that is mostly because the level-one theory of eye movements is so simple that we don't notice that it's even there. In fact, general ideas along these lines were in Gibson's thinking, I suspect, and were certainly being articulated by Marvin Minsky and Seymour Papert in the late 1960s and early 1970s. But the details to these general ideas were never filled in. In a curious sense, this was because artificial intelligence remained decerebrate. It never realized that there was a level one theory to be discovered. It remained, and often still does, stuck fast in the mud of mechanistic explanations—where memory is held to be achieved by a neural net of some kind, or by a process in a computer, or by a set of procedures.

I don't know about this. These seem quite reasonable ways of explaining memory. Why do you find them so objectionable?

Well, in simple cases like eye movements, we can think in that rather direct fashion and get away with it. But it is very dangerous to hope that this type of thinking can ever give any real insight into the computational problems that the neural mechanisms are busy solving.

For example, to take a famous and elegantly expressed case, we might discuss Minsky's frames theory a little. A frame is essentially an item to which properties may be attached. For example, consider the following properties of an elephant considered as a frame:

Name	Clyde
Color	Pink
Weight	Large
Appetite	Large

Processes can also be attached to a frame and the contents of a frame may be interconnected or indexed in various ways. In his most stimulating article, Minsky (1975) describes how many "subjectively plausible" phenomena can be thought of in this way provided that the conceptual units involved are "large" enough. But I believe the approach is fundamentally

flawed by its mechanism-based thinking. This harks back to our earlier point. If frames offered a representation and not just a mechanism, we would at once see what they are capable of representing and what they are not. This may still be done, but it has not yet been, until it has, we must be wary of ideas like frames or property lists. The reason is that it's really thinking in similes rather than about the actual thing—just as thinking in terms of different parts of the Fourier spectrum is a simile in vision for thinking about descriptions of an image at different scales. It is too impudent to be useful. Real progress can only be made in such cases by precisely formulating the information processing problems involved in the sense of our level one.

But your point isn't about just frames, is it? Doesn't it apply to almost the whole of artificial intelligence?

Yes, very true, and mechanism-based approaches are genuinely dangerous. The problem is that the goal of such studies is mimicry rather than true understanding, and these studies can easily degenerate into the writing of programs that do no more than mimic in an unenlightening way some small aspect of human performance. Weizenbaum (1976) now judges his program ELIZA to belong to this category, and I have never seen any reason to disagree. More controversially, I would also criticize on the same grounds Newell and Simon's (1972) work on production systems and some of Norman and Rumelhart's (1974) work on long-term memory.

Why, exactly?

The reason is this. If we believe that the aim of information-processing studies is to formulate and understand particular information-processing problems, then the structure of those problems is central, not the mechanisms through which their solutions are implemented. Therefore, in exploiting this fact, the first thing to do is to find problems that we can solve well, find out how to solve them, and examine our performance in the light of that understanding. The most fruitful source of such problems is operations that we perform well, fluently, and hence unconsciously, since it is difficult to see how reliability could be achieved if there was no sound, underlying method.

Unfortunately, problem-solving research has for obvious reasons tended to concentrate on problems that we understand well intellectually but perform poorly on, like mental arithmetic and cryptarithmetics* geometry-theorem proving, or the game of chess—all problems in which human skills are of doubtful quality and in which good performance seems to rest on a huge base of knowledge and expertise.

*For example, DONALD + GERALD = ROBERT. The object is to find the digit each letter stands for.

I argue that these are exceptionally good grounds for *not* yet studying how we carry out such tasks. I have no doubt that when we do mental arithmetic we are doing something well, but it is not arithmetic, and we seem far from understanding even one component of what that something is. I therefore feel we should concentrate on the simpler problems first, for there we have some hope of genuine advancement.

If one ignores this structure, one is left with unlikely looking mechanisms whose only recommendation is that they cannot do something we cannot do. Production systems seem to me to fit this description quite well. Even taken on their own terms as mechanisms, they leave a lot to be desired. As programming languages, they are poorly designed and hard to use, and I cannot believe that the human brain could possibly be burdened with such poor implementation decisions at so basic a level.

This mimicry idea—is it just the business of thinking in similes that you mentioned before?

Yes, very much so. In fact, we could draw another parallel, this time between production systems for students of problem solving and Fourier analysis for visual neurophysiologists. Simple operations on a spatial-frequency representation of an image can mimic several interesting phenomena that seem to be accomplished by our visual systems. These include the detection of repetition, certain visual illusions, the notion of separate independent channels, separation of overall shape from fine local detail, and a simple expression of size invariance. The reason why the spatial-frequency domain is ignored by image analysis is that it is virtually useless for the main job of vision—building up a description of what is there from the intensity array. The intuition that visual physiologists lack, and which is so important, is for how this may be done. As a computing mechanism, a production system exhibits several interesting ideas—the absence of explicit subroutine calls, a blackboard-like communication channel, and some notion of a short-term memory.

However, just because production systems display these side effects (as a Fourier analysis “displays” some visual illusions) does not mean that they have anything to do with what is really going on. For example, I would guess that the fact that short-term memory can act as a storage register is probably the least important of its functions. I expect that there are several “intellectual reflexes” that operate on items held there about which nothing is yet known and which will eventually be held to be the crucial things about short-term memory.

Studying our performance in close relation to production systems seems to me a waste of time, because it amounts to studying a mechanism, not a problem. Once again, the mechanisms that such research is trying to penetrate will be unraveled by studying the problems that need solving, just as vision research is progressing because it is the problem of vision that is being attacked, not neural visual mechanisms.

What about human memory? You implied that the same type of misdirection was evident there. What did you mean?

I was referring to Norman and Rumelhart's work on the way information seems to be organized in long-term memory. Again the danger is that questions are not asked in relation to a clear information-processing problem. Instead, they are asked and answers proposed in terms of mechanisms—in this case the mechanism is called an “active structural network,” and it is so simple and general as to be devoid of theoretical substance. Norman and Rumelhart may be able to say that such an “association” seems to exist, but they cannot say of what the association consists, nor do they say that to solve problem x (which we humans can solve) memory must be organized in a particular way; and that if this organization exists, certain apparent “associations” occur as side effects.

The phenomenological side of experimental psychology can do a valuable job in discovering facts that need explaining, including those about long-term memory, and the work of Shepard (1975), Rosch (1978), and Warrington (1975), for example, seems to me very successful at this; but like experimental neurophysiology, experimental psychology will not be able to explain those facts unless information-processing research has identified and solved the underlying information-processing problems, and I think that this is where we should be concentrating our energies.

What about Gunther Stent's work on the leech, though? Isn't that rather mechanism based, too?

Yes, but it is meant to be. It is concerned with elucidating the precise mechanism by which a leech swims. I value his work very highly like that of the Tübingen group's on the housefly, but I think that early hopes of generalizing very far from these results have not borne fruit, and the reason is the levels story again. What higher nervous systems must do is determined by the information-processing problems that they must solve. We may have some simple leechlike oscillators inside us, and they may, to be very farfetched, eventually help us to understand some aspects of respiration. But such results will not teach us how we see.

One has a strong urge to tie explanation to structure eventually—that, of course, was the impact of molecular biology. It has to be done here, don't you think? Or do you see the endeavor as totally hopeless?

Yes, I agree it has to be done for the central nervous system, but I doubt if it can ever be done completely. The complexity barrier is just too great. But we have started to do it, don't forget! The zero-crossing detection and directional selectivity stories are very close to neurons. Don't be too impatient about the later things! As I said earlier, I bet you could never understand the fast Fourier transform as implemented in transistors on an IBM 370. I can only understand its formulas for about 10 minutes at a

time—let alone understand a circuit diagram implementing them. One last word—I don't think that developmental and genetic programs will be able to be understood so directly in terms of underlying mechanisms. I would guess that some levels structure will eventually be needed to understand growth, because it is complicated.

Can we perhaps return to thinking rather specifically about visual perception and what actually happens when you see?

Well, are you happy with the primal sketch ideas?

I think so. The critical point seems to be that even very early vision is a highly symbolic activity. Assertions are actually made where lines end—yes, I've even accepted that terminology and am not too worried here about neurons!—and that objective lines and virtual lines are just as "real" as one another. Both can, for example, have their orientations detected and manipulated. Isn't this the idea?

Very much so. And if there is one more key idea, it is the idea of a place token and the ability to use crude selection criteria to group such tokens together and look for patterns, just as we saw in Figure 2–3.

I'm still a little unhappy about the representation of spatial relations—in the image, that is. I remember the discussion in Chapter 2 about coordinate systems, but was a little unconvinced. How can we be sure that important spatial information isn't lost?

Well, we have to be careful here, because I do not think much in the way of spatial relations is made explicit very early on. For example, certainly no intrinsic structure like the angle between two lines is. This type of information is not explicit in the full primal sketch, nor would the angle between two surfaces be in the 2½-D sketch. Such quantities do not belong to perception; their realm is that of the 3-D model representation. On the other hand, a few explicit spatial relations, like virtual lines between neighboring place tokens, often carry implicitly the entire geometry of the figure. This can be true even if the length measurements are very imprecise—perhaps only ranked by size.

A striking example of the richness of the information coming from a few clues about nearness is provided by the archaeological endeavors of Flinders Petrie. He measured the similarity of graves found along the Upper Nile by judging the number of characteristics shared by pieces of pottery found in each one. By using just this similarity information, techniques like multidimensional scaling can recover the times of burial quite accurately. The story makes fascinating reading (see Kendall, 1969), but we need note only that in two dimensions, the situation is even more constrained. I do not think there's much danger of the information being lost, but I do think only rather little spatial information is made explicit at the early stages.

So we derive the full primal sketch and then all those processes of Chapter 3 run to give us surface information? And roughly speaking, that is delivered in retinocentric polar coordinates, with perhaps slight differences for each process?

Yes, indeed, and the surface information from each process is combined in the 2½-D sketch, still in a retinocentric fashion but perhaps in a more convenient frame than the polar one. In a deep sense this is the end of pure autonomous perception. At this point the information is ready to be turned into a real 3-D model type of representation, a description that you can then remember.

I'm still unhappy about this tying-together process and the idea that from all that wealth of detail all you have left is a description. It sounds too cerebral somehow.

Well, the description can be arbitrarily rich—it's just a question of how much time and energy you spend on it. The other matter, that visual perception is just the formation of such descriptions—well, that is the conceptual leap I'm asking you to make. I personally find nothing important that this view fails to account for in general, and since we probably understand 20%–25% of the whole process already, I'm frankly ready to put my money on the rest of the process being of the same character. It's a conceptual leap, to be sure, but I think this view is worth trying to live with for a while, because thinking of visual perception in terms of the formation of particular kinds of descriptions explains so much so simply—*the structure of vision is complicated enough at the top level, and outrageously so in terms of wiring.*

And the result of those Chapter 3 processes, embodied in the 2½-D sketch, is the end of the immediate perception?

I think it's the right place to make the division, because up to here the processes can be influenced little or not at all by higher-order considerations. They deliver what they compute—no more, no less. The term *immediate perception* is a bit misleading, because these processes can take time—think of fusing a random-dot stereogram—but they do not involve scrutiny in Julesz's sense of an active intelligent examination of the image and comparison of its parts. This is compatible with the random-dot stereogram case, because we think that when the time to perceive one is long, most of the delay is due to random-walk-like movements of the eyes as they try to find somewhere to start fusion from.

If the 2½-D sketch changes every time you move your eyes, you lose it every time you move them (except possibly for small movements purely in depth). Isn't this a terribly wasteful thing to do?

It is wasteful, surely, but if you have the machinery there capable of recomputing the scene in real time, it doesn't matter that it's wasteful. In fact, it almost has to be this way, since the point of the 2½-D sketch is to assemble and represent incoming perceptual information, not to store it, and the alternative of economizing on computing power by using more memory is of no real use here. Just suppose, for example, that a 2½-D sketch had foveal resolution everywhere and was driven by a foveal retina in the usual way. Immediately, the memory has to contain out-of-date information (or nothing) in most of its capacity. This is not what the memory is for. Before resorting to almost any real storage, one must convert to something like the 3-D model representation, which is much more stable than the viewer-centered appearance of an object in a fleeting world. So the representation in which information from the different sources is assembled must be retinocentric and transient; it should have a foveal region where resolution is high, and it should reflect exactly and only what is coming in now.

These seem sensible distinctions, but they raise a difficulty I have in relating this to my own experience. The problem is that there seem to be so many different things going on in this model for perception, yet my perception has a unity, a oneness that I feel does not jibe with or at least is not reflected in these ideas. How is all the information tied together? How can one account for the unity of visual experience?

The basic idea is indeed that very many things are delivered through almost independent processes. At the 2½-D sketch level they are tied together, but only implicitly, whereas the next step is the creation of object-centered descriptions of the visible shapes (which is perhaps localized in a viewer-centered frame), and the description here is a unified object made up just by adding properties to its basic shape description, rather as a novelist adds to a description by adding qualifying adjectives.

What do you mean by being tied together "only implicitly"?

Simply that although different processes operate in different ways, there is a way of finding out when they are referring to the same visual object.

You mean if a raw primal sketch process finds an edge, and a color process finds its color, the relation between the two is implicitly available? I don't quite follow.

It's all a question of addressing. In most computers, you address information by specifying where to look for it. In some computers, you access a chunk of information by specifying pieces of the chunk. That is a content-addressable memory, and such memories are easy to build. What we might have here is a mixture of these two types of addressing—something like

"the edge at roughly position (x,y) in the visual field with an orientation within, say, 30° of some given value". That would uniquely specify the edge in question both for the raw primal sketch representation and for the output for the color processes. In this way, we can tie the two things together, at least in principle.

What, dare I ask, about all those cortical areas? Isn't it natural to expect that they should each deal with a different process?

I would not be surprised.

Then what you are hinting at is, essentially, that up to this point each process runs, perhaps in a different cortical area (by now there are 10 at least, aren't there?), and that by presenting each with rough information, which could be rough position and orientation, you define precisely which visual object you are referring to.

Yes, that is the addressing problem.

And then, in addition, you get the precise information with which that particular area or process is concerned—the particular color or disparity, for example.

Exactly. And I think that the critical point about this is that the joining together of information is done symbolically.

What do you mean by that?

It's not like adding together the three impressions that a printer uses to make a printed page of color. We never see the colors of things smudged beyond their boundaries. The point is that the rough position and orientation information is used as an address. If you want the position of an item's exact boundary, you look at the raw primal sketch. If you want its color, you look at the color process.

I see. This idea means that assembling the information must be a very active process, doesn't it? Unless something specifically notices that stereo, zero-crossing x is a brown border, these two pieces of information will remain separate.

Yes, I think one has to ask for the color of x . And we must expect much of this to go on automatically as we move our eyes around. That is what the 2½-D sketch is partly for, after all—reducing information about surface geometry from many retinocentric processes to a single, more usable, viewer-centered form. At the same time, links to descriptions of other aspects of a surface are presumably made easily accessible, in preparation for the task of constructing a three-dimensional, object-centered description.

So you think it's likely that the actual combination isn't done until the 3-D model starts being constructed?

Yes.

It's as though strings are there to all the relevant information clearly marked and labeled, but you don't pull it all together unless you start making a 3-D model.

Which may be a very coarse one or parts of a very fine one. And in the same way, one might expect other properties to be coarse (for example, greenish) or quite fine (for example, a specific shade of green).

But how does this correspond to my perceptual experience? My experience appears to be complete, not at all the halfway, ill-defined, fragmented sort of thing than you describe.

Well, first remember that our visual processes can work extremely rapidly. The time between requesting information about a part of the visual field and moving the eyes there, getting it, and linking it to a 3-D model is probably usually under half a second. The second thing is, How much of a novel scene can you recall if you look at it only very briefly? Not very much! Its coarse organization, or perhaps one or two details. And once you close your eyes, the richness is gone, isn't it? I think that the richness corresponds to what is available now, at the pure perceptual level, and what you can remember immediately is much more closely related to the 3-D model description that you create for it while your eyes are open.

I begin to see more clearly the force of the idea that perception is the construction of a description.

Yes, that is the core of the thing, and a really important point to come to terms with.

But let's suppose you're right, then, that the 2½-D sketch is retinocentric and that you compute out of it little 3-D models and hang them up in a space frame centered on you. What happens when you move your eyes a lot?

One thing is that the finely detailed shape that you were just looking at—suppose it was a porcelain cat—and for which you have just built up an elaborate description is reduced to a blob in the image when you turn your eyes to study its neighbor, a porcelain dog. If the blob can be distinguished confidently in the 2½-D sketch, then I would guess that there is a process that maintains the link between it and the 3-D model you've just finished building, so that if that blob moves, you know immediately what has moved.

But how on earth do you do that with neurons?

Hold on there—we'll face that next. But note that basically, it's not difficult computationally.

But to tie all this up with what it feels like to see—that is difficult to swallow.

It grows on you. That first step, that vision is the computation of a description, is the crucial one. Once you have accepted that, you can go on to study exactly what description and how to make it.

And again it's not at all easy for me to allow you to talk so much about computation. The brain, after all, is made of neurons, not silicon chips. But I suppose I'll get used to it. Still, if vision is the construction of descriptions, they must be implemented neurally, mustn't they? So couldn't one hope to look for neurophysiological correlates of the 2½-D sketch or of a piece of a 3-D model? That, I would find convincing.

It would be marvelous if the implementation were that simple—close to Barlow's neural dogma! My own guess is that it is more like that than a Hebb cell assembly.

There's another more general point that is still troubling me, and it has to do with the temporal continuity of perceptual experience. I understand very well how you think continuity can be held between eye movements and so forth, but this avoids the larger question of pure continuity over time. Why, if I look at a tree, do I see it continuously as the same tree? Presumably I could at any moment start a new 3-D model for it, in which case I ought to experience it as a new tree in the same spot as the old one. Yet I don't. Do you have any comments?

The permanence of the visual world—the continuity of objects in time—is an awfully important aspect of vision, and I think it's just part of our reflexes as adults that we assume it. In fact, whole aspects of processing are based on discovering and exploiting the continuity relations—the correspondence processes of Chapter 3, for example.

Another general point. You deal only with shape here. What about the recognition as being the same thing of two objects that have different shapes but the same function—like two different kinds of chair?

This theory has nothing to say about semantic recognition, object naming or function, though that is most certainly a path almost as useful as shape determination for recognition in the external world (Warrington and Taylor, 1978). I think that the problems of understanding what we mean by the semantics of an object are fascinating, but I also think that they are very difficult indeed and at present much less accessible than the problems of visual perception.

In what scene you describe is correct, would we be able to say anything about painting and drawing using this knowledge of what the visual system does with its input? Might it help to teach these skills, for example?

Perhaps, although I would hate to commit myself to a definite view yet. Nevertheless, it is interesting to think about which representations the different artists concentrate on and sometimes disrupt. The pointillists, for example, are tampering primarily with the image; the rest of the scheme is left intact, and the picture has a conventional appearance otherwise. Picasso, on the other hand, clearly disrupts most at the 3-D model level. The three-dimensionality of his figures is not realistic. An example of someone who operates primarily at the surface representation stage is a little harder—Cézanne perhaps?

With respect to other problems such as natural language, how universal is the approach you are advocating? How far can it be taken? What kind of things would it be likely to fail at?

Systems that are not modular. Things like the process by which a chain of amino acids folds to form a protein—that is to say complex, interactive systems with many influences that cannot be neglected. A burning issue in the study of natural language understanding is, of course, How modular is it, and what are the modules?

Yes, I suppose modularity is the key, but also fluency of some kind must be important, mustn't it? If a process doesn't flow well, smoothly, unattended, and without having to be patched by conscious inference, then it may have no clean theory, and that might turn it into the protein-folding class of difficult-to-understand theories. But to return to natural language, what modules have been found there?

It's not clear, and some claim it's inherently not modular and should be viewed much more heterarchically.

Doesn't that sound a little reminiscent of the early days of vision?

Yes, I'm afraid so. But there do seem to be modules and rules for modules emerging at the early level—rules for syllable formation, prosody, and most famously Chomsky's analysis of syntax.

But how much of a module is syntax? Don't artificial intelligence workers like Schank claim that syntax is not a separable module at all?

Yes, and it is clear that the syntactical decoding of a sentence cannot proceed entirely independently of its semantical analysis. But a good case is being built up that the *amount* of interaction necessary between the two is small, and the types of questions about syntax that must be answered

seem to be of a quite simple kind—for example, *snowu a paruwua tawau refer to noun phrase one or to noun phrase two?* Marcus (1980) was the first to explore these problems in detail, and he has shown that a very successful module can be made out of a parsing system. Above the level of syntax, however, few hints are currently available about what the modularity is, but I'm sure it must be present.

Why has artificial intelligence shown such resistance to traditional Chomskian approaches to syntactical analysis? Only Marcus seems to have embraced it.

I think there are two reasons. First, it is easy to construct examples in which syntax cannot be analyzed without some concurrent semantical analysis. Thus, syntax is not a truly isolated module, and this fact led the artificial intelligence people to jump to the opposite conclusion, that syntax is not a module at all. This is incorrect—the true situation seems to be that syntax is almost a module, requiring some interactions with semantics but only a very small number of types of interaction.

The second reason is our old friend, the levels. Noam Chomsky's transformational grammar is a level one theory, that is in no way concerned with *how* syntactical recognition should be implemented. It merely gives rules for stating *what* the decomposition of an arbitrary sentence should be. Chomsky's description of it as a competence theory was his way of saying this.

However, the levels idea has not been properly understood by computational linguists. Indeed, one of Winograd's reasons for rejecting Chomsky was that he could not invert the transformational structure and turn it into a parser! This observation could be made only by someone who failed to understand the distinction between levels one (what and why) and two (how). Winograd is not to be singled out for this error, however; everyone in artificial intelligence made it, and now that the linguists themselves are becoming computationally aware, they are falling into the same trap. The result is, I fear, that natural language computer programs have contributed rather little to natural language understanding, with the recent exception of Marcus (1980), who has begun to construct a genuine level-two theory of the parsing algorithm we use.

What do you feel are the most promising approaches to semantics?

Probably what I call the problem of multiple descriptions of objects and the resolution of the problems of reference that multiple descriptions introduce.

Could you expand on this?

Well, like many others in the field, I expect that at the heart of our understanding of intelligence will lie at least one and probably several

important principles about organizing and representing knowledge that in some sense capture what is important about the general nature of our intellectual abilities. While still somewhat vague, the ideas that seem to be emerging are as follows:

1. The chunks of reasoning, language, memory, and perception ought to be larger than most recent theories in psychology have allowed (Minsky 1975). They must also be very flexible, and incorporating this requirement precisely will not be easy.
2. The perception of an event or of an object must include the simultaneous computation of several different descriptions of it that capture diverse aspects of the use, purpose, or circumstances of the event or object.
3. The various descriptions referred to in point 2 include coarse ver-sions as well as fine ones. These coarse descriptions are a vital link in choosing the appropriate overall scenarios demanded by point 1 and in correctly establishing the roles played by the objects and actions that caused those scenarios to be chosen.

An example will help to make these points clear. If one reads

The fly buzzed irritatingly on the windowpane.
John picked up the newspaper.

the immediate inference is that John's intentions toward the fly are fundamentally malicious. If he had picked up the telephone, the inference would be less secure. It is generally agreed that an "insect-damaging" scenario is somehow deployed during the reading of these sentences, being suggested in its coarsest form by the fly buzzing irritatingly. Such a scenario will contain a reference to something that can squash an insect on a brittle surface—a description that a newspaper fits, but not a telephone. We might therefore conclude that when the newspaper is mentioned (or, in the case of vision, seen) not only is it described internally as a newspaper and some rough 3-D model description of its shape and axes set up, but it is also described as a light, flexible object with area. Because the second sentence might have continued "and sat down to read," the newspaper must also be described as reading matter; similarly, it must also be described as a combustible article, as something that rustles, and so forth. Since we do not usually know in advance what aspect of an object or action is important, it follows that most of the time a given object will give rise to several different coarse internal descriptions. Similarly for actions. It may be important to note that the description of fly swatting or reading or fire lighting does not have to be attached to the newspaper—a description of the newspaper is merely available that will match its role in each scenario.

Why do you think this must be so?

Because the importance of a primitive, coarse catalogue of events and objects lies in the role that such coarse descriptions play in the ultimate access and construction of perhaps exquisitely tailored specific scenarios, rather in the way that a general 3-D animal model can finish up as a very specific Cheshire cat after due interaction between the image and information stored in the catalogue of models. What existed as little more than a malicious intent toward the innocent fly after the first sentence becomes, with the additional information about the newspaper, a very specific case of fly squashing. Exactly how this is best done and exactly what descriptions should accompany different words or perceived objects is not yet known.

What about other types of processing that the brain does, such as the planning and execution of behavior? Might not these be simpler places to start looking for modules? After all, semantics is one of the most advanced areas of human ability, so it's not unreasonable to expect that it may be complex. I would try something simpler.

I think that may be excellent advice, and it reminds me of a fascinating experiment done some time ago by Stann (1969). He was running what is called a delayed-response task (see Figure 7-2). In this, a scrap of food

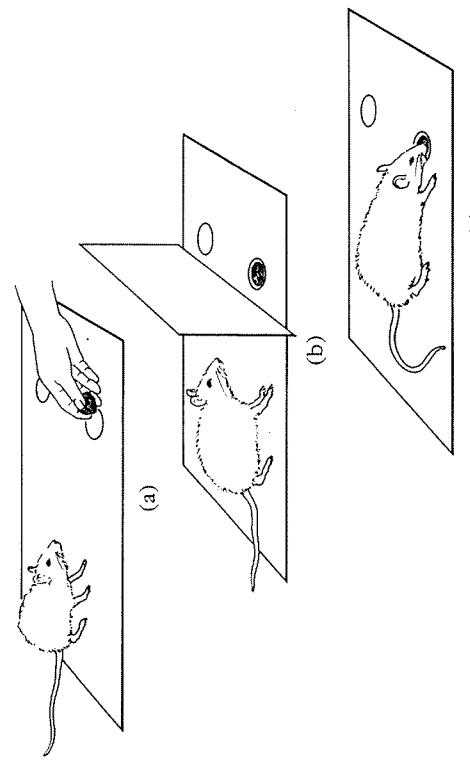


Figure 7-2. A delayed-response task. A scrap of food is placed under one of the wells in full view of the animal. Then a screen descends for a period. When the screen is raised, the animal has to choose one of the wells. If he looks under the correct one, he is rewarded with the food.

is placed in one of two wells, a screen comes down, a delay ensues, the screen lifts, and the animal is then free to choose the well in which he thinks the food is hidden. Certain portions of the prefrontal cortex are known to be involved in this task, and the animal cannot perform it if they are removed. Stamm used a technique—depolarization—whereby he could effectively disable these areas for the precise period he desired. He asked, When must the area be operating for the task to be carried out? It turned out that the animal had to have its area working as the screen came down at the beginning of the delay; if the area was knocked out at any other time, it mattered either much less or not at all!

One possible way of thinking about this experiment is this. Any real-time computer must be able to construct plans, set them up for execution under the appropriate conditions, and set the triggers for them. One cannot recompute everything afresh each time, and indeed the structure of a human personality consists in part of thousands of such little plans, all set to run a person's behavior if the appropriate conditions arise. But something must *write* these plans, and here in Stamm's experiment maybe we are seeing a simple example of this happening. As the wells are removed from view, the animal writes into its set of plans to go to the appropriate well when it can. A simple plan, but a plan nevertheless.

If we carry this idea a little further, we see that it splits the central system into what one might call the planner and the executive. The planner writes plans and their triggers to the executive, which, when the time and conditions are ripe, executes them. Is it too absurd to suggest that during hypnosis the executive becomes externally programmable and that this is why it is possible to set up plans under hypnosis that are executed later when the assigned conditions are met? The idea bears reflection, at least.

That is an interesting idea. I have not seen any previous explanation about why it should be possible to "program" someone at all, and your suggestion is certainly plausible. But what about the stereotyped nature of the programming? We are ourselves very flexible, are we not? It's a little difficult to reconcile that with a set of programmed responses.

I think that depends entirely on how large, rich, and subtle the set of responses has grown to be. If there is wide variety of responses and considerable ability to act differently in only subtly different situations, then we would be called flexible—and freer, incidentally, since we would be taking a wider range of relevant information appropriately into account. If we take no information (random response) or only one piece (compulsive response), then we are certainly not acting flexibly or freely.

That seems a sensible distinction. But as we move closer to saying the brain is a computer, I must say I do get more and more fearful about the meaning of human values.

Well, to say the brain is a computer is correct but misleading. It's really a highly specialized information-processing device—or rather, a whole lot of them. Viewing our brains as information-processing devices is not demeaning and does not negate human values. If anything, it tends to support them and may in the end help us to understand what from an information-processing view human values actually are, why they have selective value, and how they are knitted into the capacity for social mores and organization with which our genes have endowed us.

Complex cells An orientationally sensitive class of cells in the visual cortex discovered by Hubel and Wiesel. These cells are more complicated than simple cells in that their response is not a linear function of the spatial stimuli falling within their receptive fields, but they do not show any particular sensitivity to the termination of edges and bars.

Component axis A subsidiary axis of a 3-D model, for example, the neck axis in a quadruped 3-D model.

Conjunctive eye movements Eye movements that change the average direction of gaze of the two eyes.

Contour generator The locus of points on a visible surface that gives rise to a contour in the surface's image.

Convolution (*) Formally, the convolution of two functions $f(x)$ and $g(x)$ is given by $\int f(x') g(x - x') dx'$. For the case of an image, its meaning may be visualized more easily in terms of receptive fields. Suppose we place at position (x,y) in an image a weighted receptive field, perhaps with a center-surround organization. This field adds up linearly the contributions from each part of the image as "seen" through the receptive field—that is, points in the center receive a strong positive weighting and those in the surround a weaker negative weighting. The result is the value of the convolution of the image with the function represented by the receptive field weights at that one particular point (x,y) . Thus to calculate directly the convolution of the whole image, that is, for every point (x,y) , can be a computationally expensive process.

Cooperative Algorithm A nonlinear algorithm in which purely local operations appear to cooperate to produce order on a global scale in a well-regulated manner. So called after cooperative phenomena in physics, like the Ising model of ferromagnetism, superconductivity, and phase transitions in general. Cooperative algorithms share many characteristics with these phenomena.

$\nabla^2 G$ The Laplacian operator applied to a Gaussian distribution in two dimensions. The result has a Mexican-hat shape and can be written:

$$\nabla^2 G(r) = -1/\pi r^4 (1 - r^2/2\sigma^2) \exp(-r^2/2\sigma^2)$$

It is illustrated in Figure 2-9.

Depth Viewer's subjective impression of the distance to the visible surface.

Description A description is the result of applying a representation to a particular entity (see *Representation*).

Differential operators Spatial differential operators like $\partial/\partial x$ and $\partial/\partial y$ can be realized approximately by convolution operators with appropriately shaped receptive fields. Some of these are diagrammed in Figure 2-11.

Dip See *Slope*.

Glossary

Action potential The self-regenerating electrical spike that propagates down an axon, thus transmitting a signal from one cell to the next via a synapse. The mechanism of the conduction of this signal was elucidated by A. L. Hodgkin and A. F. Huxley.

Adjunct relation A flexible way of specifying the relative positions of two axes in a 3-D model, usually used to relate a component axis to the model's principal axis (see Figures 5-4 and 5-5).

Area 17 The striate cortex.

Band-pass channel A filter that allows only a particular band of frequencies to pass through it.

Bit map A convenient way of representing rough position in an image. A two-dimensional array is set in correspondence with the x - and y -coordinates in an image, and the positions of items are represented by putting a 1 at the appropriate point in the array.

Blocks world The visual domain of matte white, plane-faced blocks viewed against a dark background. Much early machine vision was conducted in this domain.

Disjunctive eye movements Eye movements that change the relative directions of gaze of the two eyes, making them more convergent or more divergent, while leaving their average direction of gaze unchanged.

Disparity If two items are positioned at different distances from the viewer, the relative positions of their images in the two eyes will differ. This difference, usually measured in minutes of arc, is called disparity. A 1-in. depth difference at a distance of 5 ft straight ahead will produce a disparity of about 1'.

Distance Usually refers to objective three-dimensional distance from the viewer to the visible surface.

DOG A function composed of the difference of two Gaussian distributions. Such functions are thought to describe the shape of the receptive fields of the retinal ganglion cells and the shape of the receptive fields associated with Wilson's four-channel model of early visual processing. They are very close in shape to the ideal function $\nabla^2 G$ (see Figure 2-16).

Eccentricity Usually refers to the angle out from the central fovea of the retina.

Emissittance angle The angle of emittance e is the angle between a ray of light reflected from a surface and the normal to the surface.

Entropy Roughly speaking, the entropy of a probability distribution measures how chaotic the distribution is. Thus the entropy is low if the distribution is concentrated around one value, and zero if it is concentrated on exactly one value. A uniform distribution has the maximum entropy. Formally, for a discrete distribution with outcomes $1, 2, \dots, i, \dots$ having probabilities $p_1, p_2, \dots, p_i, \dots$, the entropy $q(p)$ of the distribution is given by $q(p) = \sum_i -p_i \log_2 p_i$.

Fast Fourier transform A fast digital algorithm for carrying out a Fourier transform on a discrete array whose dimensions are a power of 2. It was devised by J. M. Cooley and T. W. Tukey. Recently S. Winograd devised an even faster algorithm known as the very fast Fourier transform (VFFT).

Frontal plane The plane lying perpendicular to the line of sight.

Gaussian (G) The so-called Gaussian or normal distribution has the form $G(r) = (1/2\pi\sigma^2) \exp(-r^2/2\sigma^2)$ in two dimensions.

Gradient space A way of representing three-dimensional surface orientation by a point on a two-dimensional graph, usually denoted by (p, q) (see Section 3.8 and especially Figure 3-73).

High-pass filter A filter that allows through only the high frequencies in a signal (these could be high spatial or temporal frequencies).

Horopter There are several definitions of the horopter, but in this book it refers to the zero-disparity surface for the current positions of the eyes.

Hyperacuity Humans can carry out a variety of tasks to accuracies that are more precise than the dimensions of the retinal cones from which the information originates. Foveal cones have a diameter of about 27'', yet many tasks yield accuracies of around 5'', and stereoscopic acuity may be as good as 2''. Such tasks are said to fall within the range of hyperacuity.

Incidence angle The angle of incidence i is the angle between a ray of incident light and the normal to the surface.

Ioluminance contour A reflectance map usually consists of contours of constant luminance, or isoluminance contours, plotted in (p, q) or gradient space.

Isotropic The same in all directions.

Just noticeable difference (JND) A JND experiment tests discrimination ability for a parameter over a range by measuring at each point in the range the amount the parameter has to be changed before the difference is noticed. The two test stimuli are usually juxtaposed.

Lambertian A Lambertian surface is a perfect diffuser, the reflective analogue of a blackbody radiator. Its reflectance function $\phi(i, e, g)$ is $\cos i$ and depends only on i , the angle of incidence of the illumination.

Laplacian (∇^2) Formally, $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$. It is the lowest-order isotropic differential operator.

Lateral geniculate body (LGN) The main visual nucleus between the eye and the brain. It is fed by the optic nerve, which consists of axons of the retinal ganglion cells. The axons emerging from the LGN, called the optic radiations, project to the striate cortex in the monkey and in man.

Low-pass filter A filter that allows through only the low frequencies in a signal (these could be low spatial or temporal frequencies).

Model axis An axis, associated with a 3-D model, that defines the overall extent of the shape that the model represents.

Modulation transfer function (MTF) The amplitude of the Fourier transform of a filter or function. The MTF is useful because by looking at its graph, one can tell at a glance which frequencies are passed and which are suppressed by the filter.

Occluding contour A contour in an image that is formed by an occluding edge.

Panum's area The disparity range over which stereoscopic fusion can be achieved without eye movements.

Panum's limiting case See Figure 3-19.

Phase angle The phase angle g is the angle between the incident and emitted rays.

Place token A token that marks a point of interest in an image. Such tokens have a position, and they may possess various other properties.

They are thought to be constructed during the early analysis of the spatial arrangement of an image.

Primal sketch A representation of the two-dimensional image that makes explicit the amount and disposition of the intensity changes there. The representation is hierarchical, the primitives at the lowest level representing raw intensity changes and their local geometrical structure, and those at the higher levels capturing groupings and alignments occurring among the lower items (see Figure 2–7).

Principal axis The axis of a 3-D model that most component axes adjoin, for example, the torso axis of a quadruped 3-D model.

Reflectance function Usually denoted by $\Phi(i,e,g)$, the reflectance function associated with a surface specifies what fraction of the incident light is reflected under different conditions of viewing and illumination. See Figure 3–75 and Section 3.8.

Reflectance map A graph that relates image intensities to surface orientation, not usually in a one-to-one manner. Figures 3–76 to 3–79 show some examples.

Representation A representation of a set of entities S is a formal scheme for describing them, together with rules that specify how the scheme applies to any particular one of the entities.

Retinal ganglion cells The final layer of cells in retinal processing. The axons of these cells leave the retina through the so-called blind spot and form the optic nerve.

Retinex Edwin Land's term for the processing of an image by removing all gradual changes in intensity, such as might be caused by changes in illumination, while leaving all sudden changes, such as might be due to changes in reflectance.

Rhodopsin The light-sensitive visual pigment in the rods and cones, the receptors of the eye.

Saccade A conjunctive eye movement can either be smooth or occur in a preprogrammed ballistic jump called a saccade, which takes about 160 ms to program internally. Disjunctive eye movements, on the other hand, are always smooth and are under continuous control based on feedback about the disparity between the current vergence angle and the desired vergence angle.

Shape The geometry of an object's physical surface.

Simple cells A class of orientationally sensitive cells in the striate cortex, discovered by Hubel and Wiesel and defined as simple cells by the linearity of their response to stimuli falling in their receptive fields.

Slant The angle by which a plane slants or dips away from the viewer's frontal plane. Also called *dip*.

Spatial frequency The Fourier transform of a signal that varies in time represents that signal as the sum of sine and cosine waves, each at a

different temporal frequency. If the signal varies in space rather than time, like a single image for instance, then the components by which its Fourier transform represents it are its spatial frequencies, which can be thought of as oriented sine wave gratings.

Spatial frequency channel A channel that allows only a limited range of spatial frequencies to pass through it. The early parts of the human visual system incorporate a number of spatial frequency channels, each of which is effectively less than two octaves wide; that is, the ratio of the maximum to the minimum frequency passed is less than 4 to 1.

Striate cortex The primary visual cortical receiving area in the monkey and in man. So called because of the stria of Gennari, a band of white matter running through only this region of the cortex.

Surface contour The image of a contour lying on a visible surface.

Synapse The junction between nerve cells occurring between the axon of one and the dendrite or soma (cell body) of the next. Most synapses are chemical—that is, messages are transmitted across them by release of a chemical from the axon terminal—but some synapses are electrical.

Tachistoscope A device used in psychophysical experiments for exposing the subject to brief visual stimuli.

3-D Model The basic building block of the 3-D model representation. It specifies a model axis, which defines the overall extent of the shape; the relative sizes and spatial arrangement of the (few) component axes of the model; and pointers to the shapes associated with these axes (see the boxes in Figure 5–3).

3-D Model representation An object-centered representation for shapes that includes the use of volumetric primitives of various sizes, arranged in a modular, hierarchical organization (see Figure 5–3).

Tilt The direction in which the surface slants away from the frontal plane.

2½-D Sketch A viewer-centered representation of the depth and orientation of the visible surfaces, including contours of discontinuity in these parameters (see Figure 3–12).

Vergence eye movements See *disjunctive eye movements*.

Volterra series A way of representing a certain class of nonlinear systems. Provided a function is sufficiently smooth, that is, has no discontinuities or threshold or decision points, it can be expressed as a series of polynomial terms; for example,

$$f(x,y) = ax + by + cxy + dx^2y + \dots$$

In the particular case of the flight control system of the housefly, only the lower-order terms are important.

W cells, X cells, Y cells The three classes of retinal ganglion cells. The X-cell-Y-cell distinction was originally discovered by C. Enroth-Cugell and J. D. Robson, the W cells being discovered later. These classes

have been isolated anatomically and physiologically. Y cells have the largest cell bodies, the largest receptive fields and are the least frequent (about 4% of the total ganglion cells). They have a high conduction velocity and relatively transient responses, and are subject to the shift and McIlwain effects, insensitive to color, and relatively more common in the periphery. X cells are smaller than Y cells, have smaller receptive fields, and occur more frequently than Y cells (about 60% of retinal ganglion cells are X cells). They have medium conduction velocity and relatively sustained responses, and are not so subject to the shift and McIlwain effects, often color sensitive, and relatively more common toward the fovea. W cells are very small cells with slow conduction velocities, forming perhaps 40% of the ganglion cell population. These cells, which are difficult to record from, are often directionally selective and may have other rather specific properties. Many of these cells project to the superior colliculus.

Zero-crossing Point where a function's value changes its sign.

Bibliography

- Adrian, E. D. 1928. *The Basis of Sensation*. London: Christophers. (Reprint ed. New York: Hafner, 1964).
- Adrian, E. D. 1947. *The Physical Background of Perception*. Oxford: Clarendon Press.
- Agin, G. 1972. Representation and description of curved objects. Stanford Artificial Intelligence Project Memo AIM-173. Stanford, Ca.: Stanford University.
- Anstis, S. M. 1970. Phi movement as a subtraction process. *Vision Res.* 10, 1411-1430.
- Attnave, F. 1974. Apparent movement and the what-where connection. *Psychologia* 17, 108-120.
- Attnave, F., and G. Block. 1973. Apparent motion in tridimensional space. *Percept. & Psychophys.* 13, 301-307.
- Austin, J. L. 1962. *Sense and Sensibility*. Oxford: Clarendon Press.

- Barlow, H. B. 1953. Summation and inhibition in the frog's retina. *J. Physiol. (Lond.)* 119, 69–88.
- Barlow, H. B. 1972. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1, 371–394.
- Barlow, H. B. 1978. The efficiency of detecting changes in random dot patterns. *Vision Res.* 18, 637–650.
- Barlow, H. B. 1979. Reconstructing the visual image in space and time. *Nature* 279, 189–190.
- Barlow, H. B., C. Blakemore, and J. D. Pettigrew. 1967. The neural mechanism of binocular depth discrimination. *J. Physiol. (Lond.)* 193, 327–342.
- Barlow, H. B., R. M. Hill, and W. R. Levick. 1964. Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *J. Physiol. (Lond.)* 173, 377–407.
- Barlow, H. B., W. R. Levick. 1965. The mechanism of directional selective units in rabbit's retina. *J. Physiol. (Lond.)* 178, 477–504.
- Beck, J. 1972. *Surface Color Perception*. Ithaca, N.Y.: Cornell University Press.
- Berry, R. N. 1948. Quantitative relations among vernier, real depth, and stereoscopic depth acuities. *J. Exp. Psychol.* 38, 708–721.
- Binford, T. O. 1971. Visual perception by computer. Paper presented at the IEEE Conference on Systems and Control, December 1971, Miami.
- Bishop, P. O., J. S. Coombs, and G. H. Henry. 1971. Responses to visual contours: Spatio-temporal aspects of excitation in the receptive fields of simple striate neurons. *J. Physiol. (Lond.)* 219, 625–657.
- Blomfield, S. 1973. Implicit features and stereoscopy. *Nature*, *New Biol.* 245, 256.
- Blum, H. 1973. Biological shape and visual science, part 1. *J. Theor. Biol.* 38, 205–287.
- Bouguer, P. 1757. Histoire de l'Academie Royale des Sciences, Paris; and *Traité d'Optique sur la Gradation de la Lumière* (Ouvrage posthume de M. Bouguer), l'Abbé de Lacaille, Paris, 1760.
- Braddick, O. J. 1973. The masking of apparent motion in random-dot patterns. *Vision Res.* 13, 355–369.
- Braddick, O. J. 1974. A short-range process in apparent motion. *Vision Res.* 14, 519–527.
- Braddick, O. J. 1979. Low- and high-level processes in apparent motion. *Phil. Trans. R. Soc. Lond. B* 290, 137–151.
- Brady, M. 1979. Inferring the direction of the sun from intensity values on a generalized cone. *Proc. Int. Joint Conf. Art. Intel., IJCAI-79*, 88–91.
- Braunstein, M. L. 1962. Depth perception in rotation dot patterns: Effects of numerosity and perspective. *J. Exp. Psychol.* 64, 415–420.
- Breitmeyer, B., and L. Ganz. 1977. Temporal studies with flashing gratings: Influences about human transient and sustained channels. *Vision Res.* 17, 861–865.
- Brindley, G. S. 1970. *Physiology of the Retina and Visual Pathway*. Physiological Society Monograph no. 6. London: Edwin Arnold.
- Brodatz, P. 1966. *Textures: A Photographic Album for Artists and Designers*. New York: Dover.
- Campbell, F. W. C. and J. Robson. 1968. Application of Fourier analysis to the visibility of gratings. *J. Physiol. (Lond.)* 197, 551–566.
- Campbell, F. W. C. 1977. Sometimes a biologist has to make a noise like a mathematician. *Neurosciences Res. Prog. Bull.* 15, 417–424.
- Carey, S., and R. Diamond. 1980. Maturational determination of the developmental course of face encoding. In *Biological Bases of Mental Processes*, D. Kaplan, ed., 1–7. Cambridge, Mass.: MIT Press.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.
- Chomsky, N., and H. Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8, 425–504.
- Clarke, P. G. H., I. M. L. Donaldson, and D. Whitteridge. 1976. Binocular mechanisms in cortical areas I and II of the sheep. *J. Physiol. (Lond.)* 256, 509–526.
- Clocksin, W. F. 1980. Perception of surface slant and edge labels from optical flow: A computational approach. *Perception* 9, 253–269.
- Corbin, H. H. 1942. The perception of grouping and apparent motion in visual space. *Arch. Psychol.* Whole No. 273.
- Crick, F. H. C., D. Marr, and T. Poggio. 1980. An information processing approach

- to understanding the visual cortex. In *The Cerebral Cortex*, Ed. F. O. Schmitt and F. G. Worden. (The Proceedings of the Neurosciences Research Program Colloquium held in Woods Hole, Mass., May 1979.) Cambridge, Mass.: MIT Press.
- Dev, P. 1975. Perception of depth surfaces in random-dot stereograms: A neural model. *Int. J. Man-Machine Stud.* 7, 511–528.
- DeValois, R. L. 1965. Analysis and coding of color vision in the primate visual system. *Cold Spring Harbor Symp. Quant. Biol.* 30, 567–579.
- DeValois, R. L., I. Abramov, and G. H. Jacobs. 1966. Analysis of response patterns of LGN cells. *J. Opt. Soc. Am.* 56, 966, 977.
- DeValois, R. L., I. Abramov, and W. R. Mead. 1967. Single cell analysis of wavelength discrimination at the lateral geniculate nucleus in the macaque. *J. Neurophysiol.* 30, 415–433.
- Dreher, B. and K. J. Sanderson. 1973. Receptive field analysis: Responses to moving visual contours by single lateral geniculate neurons in the cat. *J. Physiol. (Lond.)* 234, 95–118.
- Enroth-Cugel, C. and J. D. Robson. 1966. The contrast sensitivity of retinal ganglion cells of the cat. *J. Physiol. (Lond.)* 187, 517–522.
- Evans, R. M. 1974. *The Perception of Color*. New York: Wiley.
- Felton, T. B., W. Richards, and R. A. Smith, Jr. 1972. Disparity processing of spatial frequencies in man. *J. Physiol. (Lond.)* 225, 349–362.
- Fender, D., and B. Julesz. 1967. Extension of Panum's fusional area in binocularly stabilized vision. *J. Opt. Soc. Am.* 57, 819–830.
- Forbus, K. 1977. Light source effects. MIT A.I. Lab Memo 422.
- Fram, J. R., and E. S. Deutsch. 1975. On the quantitative evaluation of edge detection schemes and their comparison with human performance. *IEEE Transactions on Computers C-24*, 616–628.
- Freuder, E. C. 1974. A computer vision system for visual recognition using active knowledge. MIT A.I. Lab Tech. Rep. 345.
- Frisby, J. P., and J. L. Cloworthy. 1975. Learning to see complex random-dot stereograms. *Perception* 4, 173–178.
- Frisby, J. P., and J. E. W. Mayhew. 1979. Does visual texture discrimination precede binocular fusion? *Perception* 8, 153–156.

- Galambos, R., and H. Davis. 1943. The response of single auditory-nerve fibres to acoustic stimulation. *J. Neurophysiol.* 7, 287–303.
- Gibson, J. J. 1950. *The Perception of the Visual World*. Boston: Houghton Mifflin.
- Gibson, J. J. 1958. Visually controlled locomotion and visual orientation in animals. *Brit. J. Psych.* 49, 182–194.
- Gibson, J. J. 1966. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.
- Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Gibson, J. J., and E. J. Gibson. 1957. Continuous perceptive transformations and the perception of rigid motion. *J. Exp. Psychol.* 54, 129–138.
- Gibson, E. J., J. J. Gibson, O. W. Smith, and H. Flock. 1959. Motion parallax as a determinant of perceived depth. *J. Exp. Psychol.* 8, 40–51.
- Gibson, J. J., P. Olum, and F. Rosenblatt. 1955. Parallax and perspective during aircraft landings. *Am. J. Psychol.* 68, 372–385.
- Gilchrist, A. L. 1977. Perceived lightness depends on perceived spatial arrangement. *Science* 195, 185–187.
- Glass, L. 1969. Moire effect from random dots. *Nature* 243, 578–580.
- Glass, L., and R. Perez. 1973. Perception of random dot interference patterns. *Nature* 246, 360–362.
- Glass, L., and E. Switkes. 1976. Pattern perception in humans: Correlations which cannot be perceived. *Perception* 5, 67–72.
- Goodwin, A. W., G. H. Henry and P. O. Bishop. 1975. Direction selectivity of simple striate cells: Properties and mechanism. *J. Neurophysiol.* 38, 1500–1523.
- Gordon, D. A. 1965. Static and dynamic visual fields in human space perception. *J. Opt. Soc. Am.* 55, 1296–1303.
- Gouras, P. 1968. Identification of cone mechanisms in monkey ganglion cells. *J. Physiol. (Lond.)* 199, 533–547.
- Grant, R., and G. Svartichin. 1939. Principles and technique of the electrophysiological analysis of colour reception with the aid of microelectrodes. *Upsala Lakref Fath.* 65, 161–177.

- Green, B. F. 1961. Figure coherence in the kinetic depth effect. *J. Exp. Psychol.* 62, 272-282.
- Gregory, R. L. 1970. *The Intelligent Eye*. London: Weidenfeld & Nicolson.
- Grimson, W. E. L. 1979. Differential geometry, surface patches and convergence methods. MIT A.I. Lab. Memo 510. (Available as *From Images to Surfaces: A Computational Study of the Human Early Visual System*. Cambridge: MIT Press 1981.)
- Grimson, W. E. L. 1980. A computer implementation of a theory of human stereo vision. MIT A.I. Lab. Memo 565. *Phil. Trans. Roy. Soc. Lond. B* 292, 217-253.
- Grimson, W. E. L., and D. Marr. 1979. A computer implementation of a theory of human stereo vision. In *Proceedings of ARPA Image Understanding Workshop*, L. S. Baumann, ed., SRI, 41-45.
- Gross, C. G., C. E. Rocha-Miranda, and D. B. Bender. 1972. Visual properties of neurons in inferotemporal cortex of the macaque. *J. Neurophysiol.* 35, 96-111.
- Guzman, A. 1968. Decomposition of a visual scene into three-dimensional bodies. In *AFIPS Conf. Proc.* 33, 291-304. Washington, D.C.: Thompson.
- Harmon, L. D., and B. Julesz. 1973. Masking in visual recognition: Effects of two-dimensional filtered noise. *Science* 180, 1194-1197.
- Hartline, H. K. 1938. The response of single optic nerve fibres of the vertebrate eye to illumination of the retina. *Am. J. Physiol.* 121, 400-415.
- Hartline, H. K. 1940. The receptive fields of optic nerve fibers. *Am. J. Physiol.* 130, 690-699.
- Hasenstein, B., and W. Reichardt. 1956. Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Russelkaters. *Chlorophanus. Z. Naturf.* 11b, 513-524.
- Hay, C. J. 1966. Optical motions and space perception—An extension of Gibson's analysis. *Psychol. Rev.* 73, 550-565.
- Helmholtz, H. L. F. von. 1910. *Treatise on Physiological Optics*. Translated by J. P. Southall, 1925. New York: Dover.
- Helson, H. 1938. Fundamental principles in color vision. I. The principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination. *J. Exp. Psychol.* 23, 439-471.
- Hershberger, W. A., and J. J. Starzec. 1974. Motion parallax cues in one dimensional

- polar and parallel projections: Differential velocity and acceleration/displacement change. *J. Exp. Psychol.* 103, 717-723.
- Hildreth, E. 1980. A computer implementation of a theory of edge detection. MIT A.I. Lab. Tech. Rep. 579.
- Harai, Y., and K. Fukushima. 1978. An inference upon the neural network finding binocular correspondence. *Biol. Cybernetics*, 31, 209-217.
- Hochstein, S., and R. M. Shapley. 1976a. Linear and non-linear spatial subunits in Y cat retinal ganglion cells. *J. Physiol. (Lond.)* 262, 265-284.
- Hochstein, S., and R. M. Shapley. 1976b. Quantitative analysis of retinal ganglion cell classification. *J. Physiol. (Lond.)* 262, 237-264.
- Hollerbach, J. M. 1975. Hierarchical shape description of objects by selection and modification of prototypes. MIT A.I. Lab. Tech. Rep. 346.
- Horn, B. K. P. 1973. The Binford-Horn LINEFFINDER. MIT A.I. Lab. Memo 285.
- Horn, B. K. P. 1974. Determining lightness from an image. *Computer Graphics and Image Processing* 3, 277-299.
- Horn, B. K. P. 1975. Obtaining shape from shading information. In *The Psychology of Computer Vision*, P. H. Winston, ed., 115-155. New York: McGraw-Hill.
- Horn, B. K. P. 1977. Understanding image intensities. *Artificial Intelligence* 8, 201-231.
- Horn, B. K. P., R. J. Woodham, and W. M. Silver. 1978. Determining shape and reflectance using multiple images. MIT A.I. Lab. Memo 490.
- Hubel, D. H., and T. N. Wiesel. 1961. Integrative action in the cat's lateral geniculate body. *J. Physiol. (Lond.)* 155, 385-398.
- Hubel, D. H., and T. N. Wiesel. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* 166, 106-154.
- Hubel, D. H., and T. N. Wiesel. 1968. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol. (Lond.)* 195, 215-243.
- Hubel, D. H., and T. N. Wiesel. 1970. Cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature* 225, 41-42.
- Hueckel, M. H. 1973. An operator which recognizes edges and lines. *J. Assoc. Comput. Mach.* 20, 634-647.

- Rumman, V. A. 1971. Impossible objects as nonsense sentences. *Machine Intelligence* 6, 295–323.
- Ikeda, H., and M. J. Wright. 1972. Receptive field organization of "sustained" and "transient" retinal ganglion cells which subserve different functional roles. *J. Physiol. (Lond.)* 227, 769–800.
- Ikeda, H., and M. J. Wright. 1975. Spatial and temporal properties of "sustained" and "transient" neurons in area 27 of the cat's visual cortex. *Exp. Brain Res.* 22, 363–383.
- Ikeuchi, K. Personal communication.
- Ito, M. 1978. Recent advances in cerebellar physiology and pathology. In *Advances in Neurology*, R. A. P. Kark, R. N. Rosenberg, and L. J. Shut, eds., 59–84. New York: Raven Press.
- Ittelson, W. H. 1960. *Visual Space Perception*, New York: Springer.
- Jardine, N., and R. Sibson. 1971. *Mathematical Taxonomy*. New York: Wiley.
- Johansson, G. 1964. Perception of motion and changing form. *Scand. J. Psychol.* 5, 181–208.
- Johansson, G. 1975. Visual motion perception. *Sci. Am.* 232, 76–88.
- Johnston, I. R., G. R. White, and R. W. Cumming. 1973. The role of optical expansion patterns in locomotor control. *Am. J. Psychol.* 86, 311–324.
- Judd, D. B. 1940. Hue saturation and lightness of surface colors with chromatic illumination. *J. Opt. Soc. Am.* 30, 2–32.
- Judd, D. B. 1960. Appraisal of Land's work on two-primary color projections. *J. Opt. Soc. Am.* 50, 254–268.
- Julesz, B. 1960. Binocular depth perception of computer generated patterns. *Bell Syst. Tech. J.* 39, 1125–1162.
- Julesz, B. 1963. Towards the automation of binocular depth perception (AUTOMAP-1). In *Proceedings of the IFIPS Congress*, C. M. Popplewell, ed. Amsterdam: North Holland.
- Julesz, B. 1971. *Foundations of Cyclopean Perception*. Chicago: University of Chicago Press.
- Julesz, B. 1975. Experiments in the visual perception of texture. *Sci. Am.* 232, 34–43.
- Julesz, B., and J. J. Chang. 1976. Interaction between pools of binocular disparity detectors tuned to different disparities. *Biol. Cybernetics* 22, 107–120.
- Julesz, B., and J. E. Miller. 1975. Independent spatial-frequency-tuned channels in binocular fusion and rivalry. *Perception* 4, 125–143.
- Kaufman, L. 1964. On the nature of binocular disparity. *Am. J. Psychol.* 77, 393–402.
- Kelly, D. H. 1979. Motion and vision. II. Stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.* 69, 1340–1349.
- Kendall, D. G. 1969. Some problems and methods in statistical archaeology. *World Archaeology* 1, 68–76.
- Kidd, A. L., J. P. Frisby, and J. E. W. Mayhew. 1979. Texture contours can facilitate stereopsis by initiating appropriate vergence eye movements. *Nature* 280, 829–832.
- Koenderick, J. J., and A. J. van Doorn. 1976. Local structure of movement parallax of the plane. *J. Opt. Soc. Am.* 66, 717–723.
- Koffka, K. 1935. *Principles of Gestalt Psychology*. New York: Harcourt, Brace & World.
- Kolers, P. A. 1972. *Aspects of Motion Perception*. New York: Pergamon Press.
- Kruskal, J. B. 1964. Multidimensional scaling. *Psychometrika* 29, 1–42.
- Kuffler, S. W. 1953. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.* 16, 37–68.
- Kulikowski, J. J., and D. J. Tolhurst. 1973. Psychophysical evidence for sustained and transient detectors in human vision. *J. Physiol. (Lond.)* 232, 149–162.
- Land, E. H. 1959a. Color vision and the natural image. *Proc. Natl. Acad. Sci.* 45, 115–129, 636–645.
- Land, E. H. 1959b. Experiments in color vision. *Sci. Am.* 200, 84–94, 96–99.
- Land, E. H., and J. J. McCann. 1971. Lightness and retinex theory. *J. Opt. Soc. Am.* 61, 1–11.
- Leadbetter, M. R. 1969. On the distributions of times between events in a stationary stream of events. *J. R. Statist. Soc. B* 31, 295–302.
- Lee, D. N., 1974. Visual information during locomotion. In *Perception: Essays in Honor of James J. Gibson*, I. D. G. MacLeod and O. Pick, eds. Ithaca, N.Y.: Cornell University Press.

- Legge, G. E. 1978. Sustained and transient mechanisms in human vision: Temporal and spatial properties. *Vision Res.* 18, 69–81.
- Lettrin, J. Y., R. R. Maturana, W. S. McCulloch, and W. H. Pitts. 1959. What the frog's eye tells the frog's brain. *Proc. Inst. Rad. Eng.* 47, 1940–1951.
- Logan, B. F., Jr. 1977. Information in the zero-crossings of bandpass signals. *Bell Syst. Tech. J.* 56, 487–510.
- Longuet-Higgins, H. C., and K. Prazdny. 1980. The interpretation of moving retinal images. *Proc. R. Soc. Lond. B* 208, 385–387.
- Longuet-Higgins, M. S. 1962. The distribution of intervals between zeros of a stationary random function. *Phil. Trans. R. Soc. Lond. A* 254, 557–599.
- McCann, J. J., S. P. McKee, and T. H. Taylor. 1976. Quantitative studies in retinex theory: a comparison between theoretical predictions and observer responses to the color Mondrian experiments. *Vision Res.* 16, 445–458.
- McCulloch, W. S., and W. Pitts. 1943. A logical calculus of ideas immanent in neural nets. *Bull. Math. Biophys.* 5, 115–137.
- Mackworth, A. K. 1973. Interpreting pictures of polyhedral scenes. *Art. Intell.* 4, 121–137.
- Marcus, M. P. 1980. *A Theory of Syntactic Recognition for Natural Language*. Cambridge, Mass.: MIT Press.
- Marr, D. 1969. A theory of cerebellar cortex. *J. Physiol. (Lond.)* 202, 437–470.
- Marr, D. 1970. A theory for cerebral neocortex. *Proc. R. Soc. Lond. B* 176, 161–234.
- Marr, D. 1974a. The computation of lightness by the primate retina. *Vision Res.* 14, 1377–1388.
- Marr, D. 1974b. A note on the computation of binocular disparity in a symbolic, low-level visual processor. MIT A.I. Lab. Memo 327.
- Marr, D. 1976. Early processing of visual information. *Phil. Trans. R. Soc. Lond. B* 275, 483–524.
- Marr, D. 1977a. Analysis of occluding contour. *Proc. R. Soc. Lond. B* 197, 441–475.
- Marr, D. 1977b. Rivalrous texture stereograms. *Nature* 264, 53–56.
- Marr, D. 1978. Representing visual information. *Lectures on Mathematics in the Life Sciences* 10, 101–180. Reprinted in *Computer Vision Systems*, A. R. Hanson and E. M. Riseman, eds., 1979, 61–80. New York: Academic Press.
- Marr, D. 1980. Visual information processing: the structure and creation of visual representations. *Phil. Trans. R. Soc. Lond. B* 290, 199–218.
- Marr, D., and E. Hildreth. 1980. Theory of edge detection. *Proc. R. Soc. Lond. B* 207, 187–217.
- Marr, D., and H. K. Nishihara. 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. B* 200, 269–294.
- Marr, D., G. Palm, and T. Poggio. 1978. Analysis of a cooperative stereo algorithm. *Biol. Cybernetics* 28, 223–229.
- Marr, D., and T. Poggio. 1976. Cooperative computation of stereo disparity. *Science* 194, 283–287.
- Marr, D., and T. Poggio. 1977. From understanding computation to understanding neural circuitry. *Neurosciences Res. Prog. Bull.* 15, 470–488.
- Marr, D., and T. Poggio. 1979. A computational theory of human stereo vision. *Proc. R. Soc. Lond. B* 204, 301–328.
- Marr, D., T. Poggio, and E. Hildreth. 1980. The smallest channel in early human vision. *J. Opt. Soc. Am.* 70, 868–870.
- Marr, D., T. Poggio, and S. Ullman. 1979. Bandpass channels, zero-crossings, and early visual information processing. *J. Opt. Soc. Am.* 69, 914–916.
- Marr, D., and S. Ullman. 1981. Directional selectivity and its use in early visual processing. *Proc. R. Soc. Lond. B* 211, 151–180.
- Marroquin, J. L. 1976. Human visual perception of structure. Master's thesis, MIT.
- Maturana, H. R., and S. Frenk. 1963. Directional movement and horizontal edge detectors in pigeon retina. *Science* 142, 977–979.
- Maturana, H. R., J. Y. Lettrin, W. S. McCulloch, and W. H. Pitts. 1960. Anatomy and physiology of vision in the frog (*Rana pipiens*). *J. Gen. Physiol.* 43 (suppl. no. 2, Mechanisms of Vision), 129–171.
- Mayhew, J. E. W., and J. P. Frisby. 1976. Rivalrous texture stereograms. *Nature* 264,

- Mayhew, J. E. W., and J. P. Frisby. 1978b. Texture discrimination and Fourier analysis in human vision. *Nature* 275, 438–439.
- Mayhew, J. E. W., and J. P. Frisby. 1979. Convergent disparity discriminations in narrow-band-filtered random-dot stereograms. *Vision Res.* 19, 63–71.
- Metelli, F. 1974. The perception of transparency. *Sci. Am.* 230, 91–98.
- Miles, W. R. 1931. Movement in interpretations of the silhouette of a revolving fan. *Am. J. Psychol.* 43, 392–404.
- Minsky, M. 1975. A framework for representing knowledge. In *The Psychology of Computer Vision*, P. H. Winston, ed., 211–277. New York: McGraw-Hill.
- Mitchell, D. E. 1966. Retinal disparity and diplopia. *Vision Res.* 6, 441–451.
- Monasterio, F. M. de, and P. Gouras. 1975. Functional properties of ganglion cells of the rhesus monkey retina. *J. Physiol. (Lond.)* 251, 167–195.
- Morshon, J. A., I. D. Thompson, and D. J. Tolhurst. 1978. Spatial and temporal contrast sensitivity of neurons in areas 17 and 18 of the cat's visual cortex. *J. Physiol. (Lond.)* 283, 101–120.
- Nakayama, K., and J. M. Loomis. 1974. Optical velocity patterns, velocity sensitive neurons, and space perception: A hypothesis. *Perception* 3, 63–80.
- Narasimhan, R. 1970. Picture languages. In *Picture Language Machines*, S. Kanef, ed., 1–25. New York: Academic Press.
- Nelson, J. I. 1975. Globality and stereoscopic fusion in binocular vision. *J. Theor. Biol.* 49, 1–38.
- Neuhaus, W. 1930. Experimentelle Untersuchung der Scheinbewegung. *Arch. Ges. Psychol.* 75, 315–458.
- Newell, A., and H. A. Simon. 1972. *Human Problem Solving*. Englewood Cliffs, NJ.: Prentice-Hall.
- Newton, I. 1704. *Optics*. London.
- Nishihara, H. K. 1978. Representation of the spatial organization of three-dimensional shapes for visual recognition. Ph.D. dissertation, MIT.
- Nishihara, H. K. 1981. Reconstruction of $\nabla^2 G$ filtered images from gradients at zero-crossings. (In preparation.)

- Norman, D. A., and D. E. Rumelhart. 1974. *Explorations in Cognition*. San Francisco: W. H. Freeman and Company. See esp. 35–64.
- Pearson, D. E., C. B. Rubinstein and G. J. Spivack. 1969. Comparison of perceived color in two-primary computer generated artificial images with predictions based on the Helsen-Judd formulation. *J. Opt. Soc. Am.* 59, 644–658.
- Pettigrew, J. D., and M. Konishi. 1976. Neurons selective for orientation and binocular disparity in the visual wulst of the barn owl (*Tyto alba*). *Science* 193, 675–678.
- Poggio, G. F., and B. Fischer. 1978. Binocular interaction and depth sensitivity of striate and prestriate cortical neurons of the behaving rhesus monkey. *J. Neurophysiol.* 40, 1392–1405.
- Poggio, T., and W. Reichardt. 1976. Visual control of orientation behavior in the fly. Part II. Towards the underlying neural interactions. *Quart. Rev. Biophys.* 9, 377–438.
- Poggio, T., and V. Torre. 1978. A new approach to synaptic interactions. In *Approaches to Complex Systems*, R. Heim and G. Palm, eds. 89–115. Berlin: Springer-Verlag.
- Potter, J. 1974. The extraction and utilization of motion in scene description. Ph.D. dissertation, University of Wisconsin.
- Prazdny, K. 1980. Egomotion and relative depth from optical flow. *Biol. Cybernetics*, 36, 87–102.
- Ramachandran, V. S., and R. L. Gregory. 1978. Does colour provide an input to human motion perception? *Nature* 275, 55–56.
- Ramachandran, V. S., V. R. Madhusudhan, and T. R. Vidyasagar. 1973. Apparent movement with subjective contours. *Vision Res.* 13, 1399–1401.
- Rashbass, C., and G. Westheimer. 1961a. Disjunctive eye movements. *J. Physiol. (Lond.)* 159, 339–360.
- Rashbass, C., and G. Westheimer. 1961b. Independence of conjunctive and disjunctive eye movements. *J. Physiol. (Lond.)* 159, 361–364.
- Regan, D., K. I. Beverley, and M. Cynader. 1979. Stereoscopic subsystems for position in depth and for motion in depth. *Proc. R. Soc. Lond. B* 204, 485–501.
- Reichardt, W., and T. Poggio. 1976. Visual control of orientation behavior in the fly. Part I. A quantitative analysis. *Quart. Rev. Biophys.* 9, 311–375.

- Reichardt, W., and T. Poggio. 1979. Visual control of flight in flies. In *Recent Theoretical Developments in Neurobiology*, W. E. Reichardt, V. B. Mountcastle, and T. Poggio, eds.
- Rice, S. O. 1945. Mathematical analysis of random noise. *Bell Syst. Tech. J.* 24, 46–156.
- Richards, W. 1970. Stereopsis and stereoblindness. *Exp. Brain Res.* 10, 380–388.
- Richards, W. 1971. Anomalous stereoscopic depth perception. *J. Opt. Soc. Am.* 61, 410–414.
- Richards, W., and E. A. Parks. 1971. Model for color conversion. *J. Opt. Soc. Am.* 61, 971–976.
- Richards, W. 1977. Stereopsis with and without monocular cues. *Vision Res.* 17, 967–969.
- Richards, W., and D. Regan. 1973. A stereo field map with implications for disparity processing. *Invest. Ophthalmol.* 12, 904–909.
- Riggs, L. A., and E. W. Niehl. 1960. Eye movements recorded during convergence and divergence. *J. Opt. Soc. Am.* 50, 913–920.
- Roberts, L. G. 1965. Machine perception of three-dimensional solids. In *Optical and electro-optical information processing*, ed. J. T.ippett et al., 159–197. Cambridge, Mass.: MIT Press.
- Rock, I., and S. Ebenholtz. 1962. Stroboscopic movement based on change of phenomenal rather than retinal location. *Am. J. Psychol.* 72, 221–229.
- Rodieck, R. W., and J. Stone. 1965. Analysis of receptive fields of cat retinal ganglion cells. *J. Neurophysiol.* 28, 833–849.
- Rosch, E. 1978. Principles of categorization. In *Cognition and categorization*, E. Rosch and B. Lloyd, eds., 27–48. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Rosenfeld, A., R. A. Hummel, and S. W. Zucker. 1976. Scene labelling by relaxation operations. *IEEE Trans. Man Machine and Cybernetics SMC-6*, 420–433.
- Rosenfeld, A., and M. Thurston. 1971. Edge and curve detection for visual scene analysis. *IEEE Trans. Comput. C-20*, 562–569.
- Russell, B. 1921. *Analysis of Mind*. London: Allen & Unwin.
- Sayre, A., and J. P. Frisby. 1975. The role of monocularly conspicuous features in

- Schatz, B. R. 1977. The computation of immediate texture discrimination. MIT AI Lab Memo 426.
- Schiller, P. H., B. L. Finlay, and S. F. Volman. 1976a. Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *J. Neurophysiol.* 39, 1288–1319.
- Schiller, P. H., B. L. Finlay, and S. F. Volman. 1976b. Quantitative studies of single-cell properties in monkey striate cortex. II. Orientation specificity and ocular dominance. *J. Neurophysiol.* 39, 1320–1333.
- Shepard, R. N. 1975. Form, formation and transformation of internal representations. In *Information Processing and Cognition: The Loyola Symposium*, R. Solso, ed., 87–122. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Shepard, R. N. 1981. Psychophysical complementarity. In *Perceptual Organization*, M. Kubovy and J. R. Pomerantz, eds. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Shepard, R. N., and J. Metzler. 1971. Mental rotation of three-dimensional objects. *Science* 171, 701–703.
- Shipley, W. G., F. A. Kenney, and M. E. King. 1945. Beta-apparent movement under binocular, monocular and interocular stimulation. *Amer. J. Psychol.* 58, 545–549.
- Shirai, Y. 1973. A context-sensitive line finder for recognition of polyhedra. *Artificial Intelligence* 4, 95–120.
- Sperling, G. 1970. Binocular vision: A physical and neural theory. *Am. J. Psychol.* 83, 461–534.
- Stamm, J. S. 1969. Electrical stimulation of monkey's prefrontal cortex during delayed response performance. *J. Comp. Phys. Psychol.* 67, 535–546.
- Stevens, K. A. 1978. Computation of locally parallel structure. *Biol. Cybernetics* 29, 19–28.
- Stevens, K. A. 1979. Surface perception from local analysis of texture and contour. Ph.D. dissertation, MIT. (Available as The information content of texture gradients. *Biol. Cybernetics* 42 (1981), 95–105; also, The visual interpretation of surface contours. *Artificial Intelligence* 17 (1981), 47–74.)
- Sugie, N., and M. Suwa. 1977. A scheme for binocular depth perception suggested

- Sussman, G. J. 1975. *A Computer Model of Skill Acquisition*. New York: American Elsevier.
- Sutherland, N. S. 1979. The representation of three-dimensional objects. *Nature* 278, 395–398.
- Szentagothai, J. 1973. Synapiology of the visual cortex. In *Handbook of Sensory Physiology*, vol. 7/3B, R. Jung, ed., 269–324. Berlin: Springer-Verlag.
- Tenenenbaum, J. M., and H. G. Barrow. 1976. Experiments in interpretation-guided segmentation. Stanford Research Institute Tech. Note 123.
- Tolhurst, D. J. 1973. Separate channels for the analysis of the shape and the movement of a moving visual stimulus. *J. Physiol. (Lond.)* 231, 385–402.
- Tolhurst, D. J. 1975. Sustained and transient channels in human vision. *Vision Res.* 15, 1151–1155.
- Torre, V., and T. Poggio. 1978. A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. Lond. B* 202, 409–416.
- Trowbridge, T. S., and K. P. Reitz. 1975. Average irregularity representation of a rough surface for ray reflection. *J. Opt. Soc. Am.* 65, 531–536.
- Tyler, C. W. 1973. Stereoscopic vision: cortical limitations and a disparity scaling effect. *Science* 181, 276–278.
- Tyler, C. W., and B. Julesz. 1980. On the depth of the cyclopean retina. *Exp. Brain Res.* 40, 196–202.
- Ullman, S. 1976a. Filling-in the gaps: The shape of subjective contours and a model for their generation. *Biol. Cybernetics* 25, 1–6.
- Ullman, S. 1976b. On visual detection of light sources. *Biol. Cybernetics* 21, 205–212.
- Ullman, S. 1977. Transformability and object identity. *Percept. Psychophys.* 22, 414–415.
- Ullman, S. 1978. Two dimensionality of the correspondence process in apparent motion. *Perception* 7, 683–693.
- Ullman, S. 1979a. The interpretation of structure from motion. *Proc. R. Soc. Lond. B* 203, 405–426.
- Ullman, S. 1979b. *The Interpretation of Visual Motion*. Cambridge, Mass.: MIT Press.
- von der Heydt, R., Cs. Adorjani, P. Hanny, and G. Baumgartner. 1978. Disparity sensitivity and receptive field incongruity of units in the cat striate cortex. *Exp. Brain Res.* 31, 523–545.
- Wallach, H., and D. N. O'Connell. 1953. The kinetic depth effect. *J. Exp. Psychol.* 45, 205–217.
- Waltz, D. 1975. Understanding line drawings of scenes with shadows. In *The Psychology of Computer Vision*, P. H. Winston, ed., pp. 19–91. New York: McGraw-Hill.
- Warrington, E. K. 1975. The selective impairment of semantic memory. *Quart. J. Exp. Psychol.* 27, 635–657.
- Warrington, E. K., and A. M. Taylor. 1973. The contribution of the right parietal lobe to object recognition. *Cortex* 9, 152–164.
- Warrington, E. K., and A. M. Taylor. 1978. Two categorical stages of object recognition. *Perception* 7, 695–705.
- Watson, B. A., and I. Nachmias. 1977. Patterns of temporal interaction on the detection of gratings. *Vision Res.* 17, 893–902.
- Weissstein, N. 1973. Beyond the yellow Volkswagen detector and the grandmother cell: A general strategy for the exploration of operations in human pattern recognition. In *Contemporary Issues in Cognitive Psychology: The Loyola Symposium*, R. Solso, ed. Washington, D.C.: W. H. Winston & Sons.
- Weizenbaum, J. 1976. *Computer Thought and Human Reason*. San Francisco: W. H. Freeman and Company.
- Wertheimer, M. 1912. Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift f. Psychol.* 61, 161–265.
- Wertheimer, M. 1938. Laws of Organization in Perceptual Forms. Harcourt, Brace & Co., London. 71–88.
- Westheimer, G., and S. P. McKee. 1977. Spatial configurations for visual hyperacuity. *Vision Res.* 17, 941–947.
- Westheimer, G., and D. E. Mitchell. 1969. The sensory stimulus for disjunctive eye movements. *Vision Res.* 9, 749–755.
- White, B. W. 1962. Stimulus-conditions affecting a recently discovered stereoscopic effect. *Am. J. Psychol.* 75, 411–420.

- Williams, R. H., and D. H. Fender. 1977. The synchrony of binocular saccadic eye movements. *Vision Res.* 17, 303–306.
- Wilson, H. R. 1979. Spatiotemporal characterization of a transient mechanism in the human visual system. Unpublished manuscript.
- Wilson, H. R., and J. R. Bergen. 1979. A four mechanism model for spatial vision. *Vision Res.* 19, 19–32.
- Wilson, H. R., and S. C. Giese. 1977. Threshold visibility of frequency gradient patterns. *Vision Res.* 17, 1177–1190.
- Winograd, T. 1972. *Understanding Natural Language*. New York: Academic Press.
- Woodham, R. J. 1977. A cooperative algorithm for determining surface orientations from a single view. *Proc. Int. Joint Conf. Art. Intel., IJCAI-77*, 635–641.
- Woodham, R. J. 1978. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. *Image Understanding Systems and Industrial Applications, Proc. S.P.I.E.* 155. Also available as MIT A.I. Lab Memo 479.
- Zeeman, W. P. C., and C. O. Roelofs. 1953. Some aspects of apparent motion. *Acta Psychol.* 9, 159–181.
- Zucker, S. 1976. Relaxation labelling and the reduction of local ambiguities. University of Maryland Computer Science Center, Tech. Rep. 451.
- ABRAMOV, I., 13
Active grouping processes, 50
Acuity, 63
stereo, 155
Adaptation reflectance, 253
Adaptation studies, 10
Addition, 22
Adjunct index, 320, 321
homology information from, 322
Adjunct relation, 308–309
for image-space processor, 317
ADRIAN, E. D., 11–12
Affinity measure, 185
Afterimages, 25
Algorithms, 23–24, 25, 106–110, 330
cooperative, see Cooperative stereo
algorithm; Cooperativity
for detecting intensity changes,
54–61
for directional selectivity, 167–169
iterative methods, 201
for light-source detection, 86–88
for locally parallel structure, 81–83
- for motion correspondence, 196–199
and psychophysics, 26
for raw primal sketch, 70–72
of retinex theory, 255–256
for stereo matching, 116–124; *see also*
Matching process
for transformational grammar, 28
for velocity field, 177–179
Ambiguous matches, 141
Ames room illusion, 141
Amplifiers, 11–12
AND gate, 66
and moving zero-crossing, 175
AND-NOT gate, 108–110
and directional selectivity, 163–165
Angle of emittance, 244
Angle of incidence, 244
Animal shapes, 299
ANSTRIS, S. M., 179
Aperture problem, 165–166
Apparent motion, 182–215
correspondence problem, 164,
184–187, 188–205

- Apparent motion (*continued*)
 determinants of, 160
 fusion of, 202–203
 optical flow, 212–215
 splitting of, 202–203
- Arabic numeral system, 20, 21
- Artificial intelligence, 344, 346–347
 and computational theory, 28
 and syntactical analysis, 357
- Artificial Intelligence Laboratory, MIT, 15
- Assertions, 350
- Associative property of addition, 22, 23
- ATTNEAVE, F., 196, 205, 327
- AUSTIN, J. L., 4, 30, 31, 309
- AUTOMAP program, 122, 124
- Axes
 based primitives, 304–305
 principal axis, see Principal axis
 slants, 318, 323
- Band-limited function, 289
- Band-pass filters
 for image, 75
- BARLOW, H. B., 4, 12–14, 16, 26, 97, 110,
 152, 154, 155, 162–163, 167, 262, 336,
 340, 355
- BARROW, H. G., 35, 271
- Bars, 51
- Bar and raw primal sketch, 72–73
- BATAHLI, J., 177–181
- BECK, J., 88, 89
- BENDER, D. B., 14
- Bernoulli ring, 250, 251, 257, 260
- Bergen, J. R., 10, 62, 170
- Bernoulli's equation, 345
- BERRY, R. M., 148
- BEVERLEY, K. L., 182
- Binarized representation, 59
- Binary map, 71, 72–73
- Binary numeral system, 20, 21
- Binford-Horn line finder, 17
- Bishop, P. O., 14, 167
- Bistable neural network, 25–26
- Bit map, 79–80
- BLAKEMORE, C., 13, 152, 154
- Blindness, stereo, 125–126
- Blobs, 51–52
- detectors, 343
 and raw primal sketch, 72–73
- Block, G., 196
- Blocks world, 17–18
- BLOOMFIELD, S., 291
- BUTTM, H., 304–305

- of surface interpolation, 288–291
 of velocity field, 175–177
- Computer vision, 16–19
 blocks world, 17–18
 edge detection, 16–17
 image segmentation, 270–271
 recognition, 35–36
 shading and photometric stereo, 239–250
 stereo matching, 122
 cone, generalized, 223–224, 225, 305, 314
- Consistency
 of object's identity, 204
- Constituent elements, motion of, 190–191
- Constraints, 22–23, 103–105, 267
- on addition, 22, 23
 and directional selectivity, 176–177
 rigidity constraint, 209–211
 and shading, 248
 for shape from occluding contours,
 219–223
- Continuity, 355
 of spatial localization, 68–70
 for surface orientation discontinuities,
 225
 and stereo disparity, 112–114
 and 3-D model, 321–325
- Continuation process, 287
- Chessboard pattern, 76–77
- CHANG, J. J., 122, 125, 127, 141, 143, 144
- Channels, 62–63, 337
- models for structure of, 135
- time derivative, 170–172
- Chessboard pattern, 76–77
- CHOMSKY, N., 28–29, 344, 356–357
- CLARKE, P. G. H., 125, 154
- CLATWORTHY, J. I., 125, 150, 151
- CLOCKSMITH, W. F., 212
- Collinearity of tokens, 49
- Color, 250–264
- Helson-Judd formulation, 252–253
 nonlinear intensity changes, 261–262
 and raw primal sketch, 352–353
 retinex theory of, 253–258
 and simultaneous contrast, 259–261
 trichromatic image measurements,
 262–264
- Commutative property of addition, 22, 23
- Computational constraint, 114
 as matching condition, 115
- Computational theory, 23, 25, 27–29,
 103–104, 330
- of directional selectivity, 165–167
 of early vision, 44–51
 independence of, 337, 339
 for independently moving surfaces,
 175–177
 of motion correspondence, 196–199
 retinex theory, 253–255
 of stereo vision, 111–116, 155–156
 of structure from motion, 209–210
- minimal mapping theory, 196–199
 strength (CS), 196–199
 and surface contours, 231–233
 tokens, 188
 two-dimensionality of, 193–196
- Cortical areas, 353
- Counterrotating cylinders, 182–183,
 205–206
- Craik-Cornsweet illusion, 257
- CRICK, F. H. C., 155, 335
- Cryptarithmetic, 347
- CS. See Correspondence strength
- CUMMING, R. W., 214
- CYNADER, M., 182
- DAVIS, H., 12
- Decomposition, 314
- of description, 305–307
- Delayed response task, 359–360
- Density gradient, 255–258
- Depolarization, 360
- Depth
 brightness versus, 215
 and disparities, 144–147, 208
 linear interpolation in, 290–291
 representation of, 281–282
 and 2½-D sketch, 277
- Description, 20–21
- levels of, 24–27
- Detectors, 16, 31, 336, 341, 342–344
- DEUTSCH, E. S., 16
- DEV, P., 122
- Dev's algorithm, 122–123
- DeVAOIS, R. L., 13, 258
- Diamond, R., 310
- Difference of Two Gaussians. See DOG
- Diffraction, 182
- DiP. See Slants
- Diplopia, 145
- Diamond, R., 310
- Directional selectivity, 159–182
- algorithm for, 167–169
 and AND-NOT gate, 163–164
 computational theory of, 165–167
 constraints on, 176–177
- Discontinuities
 and independently moving surfaces,
 175–182
- loom, 182
- model, 110
- neural implementation for, 169–175
- Discontinuities
 continuity of, 49–50
 and independently moving surfaces,
 175
- and interpolation, 289
- motion, direction of, 51

- Discontinuities, (*continued*)
and raw primal sketch, 71, 72–73
and spatial organization, 23–94
surface orientation, 225–226
and terminations, 78–79
in 2½-D sketch, 277, 288
- Disjunctive eye movements, 150
- Disparity, 100
and depth, 144–147, 208
and distance, 155–156
neural implementation of, 152–155
sensitivity, 152–155
stereo disparity, 111–116
and geometry, 81
and stereopsis, 279–280
and texture, 238
- DOG (difference of two Gaussians), 62, 63, 64
- for channel structures, 135
- DONALDSON, I. M. I., 125, 154
- DREHER, B., 172
- Drifting grating, 287
- EBCHEK, B., 125, 154
- Flight example, 345
- Flow, continuity of, 50–51
- Fly, flight-control system of, 32–34
- Forbidden zones, 176–177
- Forasus, K., 88, 89
- Foreshortened axis, 317
- Formal scheme, 21
- Four-points, three views, 188
- Fourier transform, 23, 76–77, 342
fast (FFT), 364
herringbone pattern, 46
- Fovea, 284
- FRAZER, J. R., 16
- Framed motion. See Apparent motion
- Frames theory, 346–347
- FRENK, H. R., 163
- FRÉUDER, E. C., 35, 36, 271
- FRISBY, J. P., 76, 97, 111, 125, 146, 147, 150, 151, 203
- Frog retina experiment, 12–13
- FUKUSHIMA, K., 122, 124
- Fusion, 144
- of apparent motion, 202–203
and correspondence strength, 196
neural implementation of, 152–155
- GALAMBOS, R., 12
- Ganglion cells in frog retina, 12–13
- GANZ, L., 170
- Gaussian filter, 56
- and interval distributions for zero-crossings, 136–137
and zero-crossings, 54, 56, 136–137
- Generalized cone, 223–224, 225, 304, 314
- Geometry
hypothesis of geometrical origin for perceptual texture boundaries, 94–95
image properties for, 80–81
and intensity, 41
and primitives, 93
- reflectance distinguished, 43
- Gestalt theory, 8
- grouping by common fate, 209
- law of Prägnanz, 187
- and motion, 159
- and similarity pairings, 186–187
- and motion, 159
- Feature detector, 12–13, 16, 31, 336, 341, 342–344
- FENTON, T. B., 144
- FENDER, D., 85, 125–129, 144, 150, 280
- Fields in images, 8
- Filling-in problem, 285–287
- Fink-Heimer technique, 14
- FINLAY, B. L., 14, 175
- First-order directional derivatives, 56–57
- Fitts' law, 12–14, 16, 49, 155, 154, 155, 175, 340, 343, 363
- Flux, 29–30
- Inverse property of addition, 22, 23
- Ising model of ferromagnetism, 122
- Glossiness, 89
- GOODWIN, A. W., 167
- GORDON, D. A., 212
- GOURAS, P., 13, 262
- Graceful degradation principle, 106
- Gradient space, 240–243
- coordinates, 27
- Grammatical approach, 28–29, 344
- Grandmother cell, 15
- GRANT, R., 12
- Granule cells, 155
- Grassfire technique, 304–305
- Gratings, sensitivity to, 61–62
- Gray-level images, 188–189
- GREEN, B. F., 209
- GREGORY, R. L., 271
- GRIMSON, E., 111, 130, 140, 147, 285–286, 290–291
- GROSS, C. G., 13–14
- Grouping processes, 91–98
by common fate, 209
and correspondence process, 199
- GUTZMAN, A., 17, 344
- HACKER program, 342
- Hand-detectors, 14
- and grandmother cell, 15
- Hardwired implementation, 25
- Hardwired table of similarities, 185
- HARMON, L. D., 19, 73–74
- HARTLINE, H. K., 12
- HASSENSTEIN, B., 110
- HAY, C. J., 209
- Hebb cell assembly, 555
- HELMHOLTZ, H. L. F. von, 8, 207
- HELSON, H., 250–253, 260
- Helson-Judd formulation, 252–253, 257
- HENRY, G. H., 14, 167
- Herring grid, 250, 251
- Herringbone pattern, 46, 76
- HERSCHBERGER, W. A., 207
- HUDDRETH, E., 52, 54–57, 63, 68, 70, 111, 135, 152
- HUL, R. M., 12
- HINTON, G., 312
- HIRAI, Y., 122, 124
- HOCHSTEIN, S., 172
- HODGKIN, A. L., 362
- HOLLBRACH, J. M., 303, 310
- Homology information, 322
- HORN, B. K. P., 16–19, 226, 239–240, 245, 248, 249, 255, 258, 269, 290, 339
- Housely, flight control system of, 32–34
- HUBEL, D. H., 12–14, 16, 49, 155, 154, 155, 175, 340, 343, 363
- HUECKE, M. H., 16
- Glass patterns, 47–48, 81–86
- algorithm for finding, 107

- HUMMEL, R. A., 271
- HUXLEY, A. F., 362
- Hyperacuity, 155
- Hypothesis of geometrical origin for perceptual texture boundaries, 94–95
- Hysteresis, 126–127
- and memory buffer, 129
- Identity of object, consistency of, 204
- IKEDA, A., 164
- Illuminants, 254–255
- Illumination, 90–96
- conditions of, 43
- and intensity, 41
- and shading, 243–245
- Ullman's studies, 86–88
- Illusions, 30
- Image-space processor, 317–318
- Images, 31
- and edge detection, 16
- initial filtering of, 338
- intensity, see Intensity
- line drawings, 17
- resolution, 43
- segmentation, 270–271
- shape information from, 37–38
- spatial arrangement of, 79–86
- and 3-D model, 322
- Immediate perception, 351
- Implementation, 330
- Incidence, angle of, 244
- Independent fusion, 146
- Independently moving surface, 175–182
- algorithms for, 177–179
- Indexes, 318–321; see also Adjunct index
- Parent index
- Information-processing tasks, 3–4
- Intensity, 16
- and blurring images, 56
- factors of, 41
- and Laplacian operator, 57–59
- and shading, 243–245
- and shadows, 89
- zero-crossings, see Zero-crossings
- Interpolation process, 285–287
- computational theories, 288–291
- Inertialstabilis interval (ISI), 161
- Invariants, 29–30
- Inverse property of addition, 22, 23
- Ising model of ferromagnetism, 122

- LIU, M., 14
 ITTELSON, W. H., 275
JACOBS, G. H., 13
 JAMES, R. C., 101
 JARDINE, N., 47, 340
 JOHANSSON, G., 209
 JOHNSTON, I. R., 214
 JUDG, D. B., 252–253, 260
 JULESZ, B., 9, 73, 78, 85, 96–97, 101–102,
 112, 115, 122, 124–129, 141,
 143–144, 146, 159–160, 204,
 280–281, 287, 291
 Kanizsa triangle, 251
 KAUFMAN, L., 125
 KELLY, D. H., 164
 KENDALL, D. G., 350
 KIDD, A. L., 97, 146
 KOENDERINK, J. J., 214
 KOFFKA, K., 159, 187
 KOLERS, P. A., 196
 KONISHI, M., 14, 154
 KRUSKAL, J. B., 340
 KUFFLER, S. W., 12, 14, 340
 Kulikowski, J., 164
 Lambertian, 245
 Lambertian surfaces, 245–247
 LAND, E. H., 17, 86, 252, 253–258, 289
 Laplacian operator, 54, 57–59
 and linear interpolation, 290–291
 LASNIK, H., 29
 Lateral geniculate cells, 64, 258
 LEADBETTER, M. R., 135
 Least common multiple principle, 106
 LEE, D. N., 212
 Leech study, 349
 LEGGE, G. E., 170
 LETTIN, J. Y., 340
 Levels of explanation, 19–20, 22–29, 103,
 337
 LEVICK, W. R., 12, 110, 163, 167
 Lightness, 250–264
 and discontinuities, 289
 retinex theory of, 253–258
 and simultaneous contrast, 259–261
 Light source. *See* Illumination
 Linear interpolation, 290–291
 Line drawings, 17
 and rotation, 10–11
 Linguistics, 28–29
 LIS programming language, 28, 342
 Local orientation, 81–83
 LOGAN, B. F., 67
 Logan's theorem, 67
 Long-term memory, 347, 349
 LONGUET-HIGGINS, M. S., 135
 LONGUET-HIGGINS, H. C., 188, 212, 214
 Looming, 182
 LOOMIS, J. M., 212
 MCCANN, J. J., 17, 86, 252, 253–258, 289
 McCULLOCH, W. S., 110
 MACH, E., 259, 327
 Mach vision. *See* Computer vision
 MCKEE, S. P., 155
 MACKWORTH, A. K., 17, 240, 344
 MADHUSUDHAN, V. R., 203
 MARCUS, M. P., 28, 357
 MARROQUIN, J. L., 49, 50
 Mass-action ideas, 339–340
 Matchable feature pairs, 127
 Matching algorithm, 130–132
 and depth, 144–147
 formulation of, 138–140
 Matching process, 115–116
 and false targets, 132–133
 input representation for, 130
 looming, 182
 and Panum's fusional area, 144
 uniqueness property, 140–144
 MATURANA, H. R., 12, 163
 MAYHEW, J. E. W., 76, 97, 111, 125, 203
 MEAD, W. R., 13
 Memory, 346, 347, 349
 Memory buffer, 129
 Mental rotation, 10–11
 METTELLI, F., 89
 Methodology of approach, 331
 METZLER, J., 10–11
 MILES, W. R., 9, 102, 159, 182
 MILLER, J. E., 125, 126, 146
 Mimicry phase, 345, 348
 Minimal mapping theory, 196–199
 MINSKY, M., 15–16, 270, 300, 346, 358
 Miniworld, 344–345
 MIT, Artificial Intelligence Laboratory, 15
 MITCHELL, D. E., 144, 150
 Modular decomposition, 305–307
 Modular design principle, 102, 325,
 356–357
 Modulation transfer function, 170
 Modules, 102
 MONASTERIO, F. M. DE, 262
 MONDRIAN, P., 86, 87
 Mondrian stimuli, 86–87
 Mondrians, 253–255
 Monkey, disparity detector in, 154
 Monocular projection, 9
 Motion, 50–51, 159–175; *see also*
 Apparent motion; Directional selectivity
 Motor system, 14–15
 MOYSHON, J. A., 164
 Muller-Lyer figure, 291
 Off-center cells, 64, 65
 and electrophysiological recordings,
 173
 OUM, P., 212
 Oncenter cells, 64, 65
 and electrophysiological recordings,
 173
 Operators
 and edge detection, 16
 and illumination, 88–89
 Laplacian operator, 54, 57–59
 Optical flow, 212–215
 Organization of shape information, 302,
 305–307
 Orientation. *See* Surface orientation
 Orientation-and-depth map, 129
 Oriented-activity neuron, 80
 Owl, disparity detectors in, 154–155
 PALM, G., 122, 271
 Panum's fusional area, 135, 144
 Panum's limiting case, 202
 illustration of, 142
 and uniqueness, 141
 PAPERT, S., 346
 Parallel surface contours, 230–231
 Parent index, 320, 321
 Homology information from, 322
 Parietal lesions, 35
 PARKS, E. A., 252
 Parsing algorithm, 357
 PEARSON, D. E., 252
 Pentose triangle, 294
 Perspective projection, 211–212
 PETRIE, F., 350
 PERTIGREW, J. D., 13, 14, 152, 154
 Philosophy of perception, 3–7
 Photometric stereo, 249–250
 PICASSO, 217, 218–219, 356
 Pick-up block idea, 341–342
 Picture element, 31
 PITTS, W., 110
 Pixel, 31
 PLANNER system, 342
 Poggendorff figure, 291
 POOGILO, T., 19, 26, 32, 34, 52, 63, 67, 85,
 102, 107, 108–111, 113, 116–117,
 122, 124, 125, 130, 146, 152, 154, 155,
 269, 271, 335, 343
 Pointillism, 356
 POTTER, J., 209

- PRASDHY, K., 159, 188, 212, 214
 Primal sketch, 37, 42, 52, 330, 350–351
 and correspondence, 189–193
 higher primitives in, 91–93
 illustration of, 53
 raw primal sketch, *see Raw primal sketch*
 Primitives, 51–52, 330
 and correspondence problem, 185
 extensions of, 310
 higher primitives, 91–93
 of raw primal sketch, 71
 of shape representation, 300–302
 Principal axis, 307–308, 316–317
 Processes, 22–24, 330–331
 grouping of, 91–98
 implicit assumptions for, 267
 information from images, 276
 and modules, 102
 and representations, 332
 segmentation, 270–271
 table of, 265
 Program control structures, 271
 Property lists, 28, 342
 Psychophysics, 9–11, 14–15, 26–27,
 101–102, 105
 of brightness, lightness, and color, 251,
 254, 260, 262
 of early vision, 61–64
 of the fly, 32–34
 of motion, 159–162, 164, 166–167, 170,
 179, 182, 189–196, 200, 206–209
 of shape recognition, 325, 328
 of stereo vision, 111, 125–128, 141–144,
 146–148, 150–151, 280–282, 286–287
 of texture discrimination, 96–98
 Pulling effect, 144
 Pure sine waves, 133
 Purkinje cells, 14
 RAMACHANDRAN, V. S., 203
 Random-dot stereograms. *See*
 Stereograms
 RASHBASS, C., 128, 150
 Raw primal sketch, 68–74
 and color, 352–353
 Real shape concept, 31
 Receptive field of cell, 12
 Recognition. *See* Shape recognition
 Recognize, segment and, 35–36
 Reductionist approach, 14
 Reflectance, 17
 adaptation reflectance, 253
 and color, 250
 function, 244
- geometry distinguished, 43
 and glossiness, 89
 and intensity, 41
 and shading, 243–245
 Reflectance map
 and photometric stereo, 249–250
 and shading, 245–247
 REGAN, D., 125, 182
 REICHARDT, W., 32, 34, 110
 REITZ, K. P., 43
 Relaxation process, 323
 Representation, 20–21, 25
 of depth, 281–282
 for early vision, 51–53, 71–73
 property list for, 342
 input representation, *see* Input representation
 for process, 23, 332
 and psychophysics, 26, 27
 sequence of, 36
 shape representation, 272, 295–328
 spatial arrangement of image, 80–81
 tokens, 44
 2½-D sketch, 268–269, 275–294
 uses of, 32
- Representational theories of mind, 6
 Resonating to invariants, 29–30
 Retinal ganglion cells. *See X cells; Y cells*
 Retinex theory, 17, 253–258
 Retinocentric frame, 42, 283–284
 Reversed phi phenomenon, 179
 RICE, S. O., 135
 RICHARDS, W., 125, 144, 217, 252
 RIGGS, L. A., 128, 150
 Rigidity assumption, 210–211
 Rigidity constraint, 209–211
 RULE, M., 94, 95, 191, 193, 199
 Rites of Spring, 217, 218–219
 ROBERTS, L. G., 17, 321
 ROBSON, J., 10, 61–62, 66, 172
 ROCCHA-MIRANDA, C. E., 14
 ROCK, I., 196
 RODECK, R. W., 64, 65, 172
 Roman numeral system, 20, 21
 ROSCH, E., 27, 349
 ROSENBLATT, F., 212
 ROSENFIELD, A., 16, 271
 Rotation and three-dimensional objects,
 10–11
 ROBENSTEIN, C. B., 252
 RUMELHART, D. E., 347, 349
 RUSSELL, B., 344
- ROUSE, J., 172
 ROUSE, W. M., 249
 Similarity, 47–49
 Silhouettes, 217, 218–219
 and apparent motion, 185–186
 and Gestalt movement, 186–187
 and distance, 279–280
 grassfire technique, 304–305
 SIMON, H. A., 347
 Simple cells, 66, 167–169, 174
 Simultaneous contrast, 259–261
 Sine waves, 133
 Size and shape representation, 301–302
 Sketch, 42
 Slants, 27, 234
 for axes, 323
 and texture, 234, 235–238
 SMITH, R. A., JR., 144
 Smoothness

- SANDERSON, K. J., 472
 SAYE, A., 125, 150
 SCHATZ, B. R., 86, 97
 SCHILLER, P. H., 14, 175
 Secondary illumination, 243
 Second-order directional derivatives, 56–57
 Segment and recognize, 35–36
 Segmentation, 270–271, 336
 Semantics of objects, 355–359
Sense and Sensibility, 30
 Sense atoms, 340
 Sense-data, 30
 Sensitivity and shape representation, 298
 Sequence of representations, 36
 Shading, 239–250
 and photometric stereo, 249–250
 shape from, 19
 Shadow boundaries, 88–89
 and transparency, 90
 Shape contours, 215–233; *see also*
 Contour generators
 occluding, *see* Occluding contours
 and reflectance map, 246–247
 surface contours, 226–233, 278
 Shape from shading, 19, 239–250
 Shape information, 37–38
 Shape recognition, 295–328
 and derivation, 321–325
 requirements for, 326
 Shape representation, 272, 295–328
 SHAPLEY, R. M., 172
 Sheep, disparity detectors in, 154
 SHEPARD, R. M., 10–11, 27, 349
 SHIRAI, Y., 17
 Short-range system, 160–162, 166–167
 SHIBSON, R., 47, 340
 Silhouettes, 217, 218–219
 and generalized cone, 224
 grassfire technique, 304–305
 SILVER, W. M., 249
 Similarity, 47–49
 cooperative algorithms for, 116–124
 and discontinuity, 290
 and distance, 279–280
 biological evidence, 125–127
 computational theory, 336
 constraints, 112–114
 cooperative algorithms for, 116–124
 matching process, *see* Matching process
 Stereo vision, 100
 Stevens' algorithm, 82–83
 STEVENS, K. A., 27, 47, 81–82, 84–86, 107,
 226–233, 234–238, 344
 Stick figure representation, 299,
 304–305
 STONE, J., 64, 65, 172

- Structure from motion, 184, 187–188, 205–212
and optical flow, 214
rigidity constraint, 209–211
Subjective contours, 50, 51
and apparent motion, 203
and virtual lines, 81
- Sugie, N., 122, 123
- Surface-based primitives, 301
- Surface contours, 226–233, 278
- Surface geometry changes, 276
- Surface orientation, 27
discontinuities, 225–226
and disparity change, 156–159
gradient space of, 240–243
linear interpolation in, 290–291
and photometric stereo, 249–250
and 2½D sketch, 277
- Surfaces, 44
contours, 226–233, 278
hierarchical organization, 44–45
texture, see Texture
- Sussman, G. J., 342
- Sutherland, N. S., 326
- Suwa, M., 122, 123
- Svartichin, G., 12
- Switkes, E., 47, 84
- Symmetry, 310
Syntax, analysis of, 356–357
- Szentagothai, J., 14, 201
- Tachistoscopic presentation of images, 150
- Taylor, A. M., 35, 328, 355
- Tennenbaum, J. M., 35, 271
- Terminations, 76–79
of bars, 51
and correspondence, 189–191
- Tecture, 233–239
boundaries, 94–95, 97
discrimination, 96–98
isolation of elements, 234–238
and stereopsis, 146–147
- Thompson I. D., 164
- 3-D model, 37, 302–309, 330, 354–356
catalogue of, 318–321, 322
extensions of, 309–313
3-D model description, 306–307
derivation for, 313–317
Threshold
detection studies, 10
and retinex theory, 258
- Thurston, M., 16
- Tilt and texture, 234, 235–238
Time derivatives, 169–170
for channels, 170–172
- Tokens, 44
boundaries, 93–96
collinearity of, 49
correspondence tokens, 188
and minimal mapping theory, 197
Place tokens, see Place tokens
and stereo disparity, 114
Tothurst, D. J., 64, 164, 170
Top-down school of thought, 100–101
- Toore, V., 108–110, 152
- Trace theory of grammar, 29
- Transformational grammar theory, 28–29, 344
- Transparency, 89–90
- Trichromatism, 9
- image measurements, 262–264
- Trigger features, 12
- Trowbridge, T. S., 43
- Tukey, T. W., 364
- Two-color projection demonstration, 252
- 2½D sketch, 37, 42, 268–294, 330, 350–352
- coordinate systems for, 283–285
discontinuities in, 288
general form of, 277–279
orientation and depth, 129
representation forms, 279–283
and segmentation, 270
and subjective contours, 81
and vergence movements, 149–151
- Tuer, C. W., 43, 146, 290
- Ullman, S., 52, 67, 86–88, 97, 102, 106, 111, 152, 159, 161, 162, 164–165, 167, 174–175, 179, 182–183, 185, 187, 188, 189–192, 193–201, 204–206, 208–211, 214–215, 288, 343
Unconventional views, 328
Uniqueness
and matching, 115, 140–144
and shape recognition, 297
- Van Doorn, A. J., 214
- Vatan, P., 314
- Vectors, three-dimensional, 308–309
- Velocity field, 51, 175–176, 179
and minimal mapping theory, 197–198
- Vergence eye movements, 128
and 2½D sketch, 149–151
Very fast Fourier transform, 364
- VidyaSagar, T. R., 203
- Viewer-centered coordinates, 317–318
- Viewpoint and intensity, 41
- Virtual line, 81–86
- Vividness, perceptual, 282
- Volman, S. E., 14, 175
- Volterra series expansion, 34
- Volumetric primitives, 301
- Von der Heydt, R., 154
- WALLACH, H., 9, 102, 105, 159, 182, 209
- WALTZ, D., 17, 225, 291, 344–345
- WARRINGTON, E., 27, 35, 36, 328, 349, 355
- WATSON, B. A., 170
- WEISSTEIN, N., 287
- WEIZENBAUM, J., 347
- WERTHEIMER, M., 8, 159
- WESTHEIMER, G., 128, 150, 155
- Wheel, rotation of, 189, 190
- White, 252–253
- WHITE, B. W., 285
- WHITE, G. R., 214
- WHITERRIDGE, D., 125, 154
- WIESE, T. N., 12–14, 16, 49, 135, 154, 175, 340, 343, 363
- WILLIAMS, R. H., 150
- WILSON, H., 10, 62–64, 135, 170, 179
- WINOGRAD, T., 28, 341, 342, 344, 357, 364
- WITKIN, A., 250, 287
- WOODHAM, R. J., 248, 249
- WRIGHT, M. J., 164

VISION

A Computational Investigation into the Human Representation
and Processing of Visual Information

DAVID MARR

Late of Massachusetts Institute of Technology

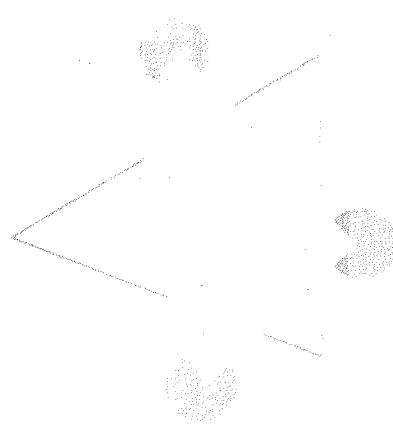
"*Vision* will be seen as a milestone in the history of the subject. Even if no single one of Marr's detailed hypotheses ultimately survives, which is unlikely, the questions he raises can no longer be ignored and the methodology he proposes seems to be the only one that has any hope of illuminating the bewildering circuitry of the central nervous system. David Marr's lifework will have been vindicated when neuroscientists cannot understand how it was ever possible to doubt the validity of his theoretical maxims." —*Science*

Presenting a revolutionary new concept of visual perception, *Vision* takes as its central tenet the notion that one cannot understand what seeing is and how it works unless one understands the underlying information-processing tasks being solved. This level of understanding—the "computational theory of a process"—identifies specific problems while exploring logical strategies and assumptions upon which solutions may be based. *Vision* is a nontechnical explication of the computational theory developed by the author and his colleagues at M.I.T. It focuses on the more abstract aspects of visual information processing, showing what is involved in solving stereo vision, the different ways motion may be used to recover structure from a sequence of images, and how the results of such processes can be represented efficiently for recognition.

The new level of understanding outlined in *Vision* complements such traditional disciplines as psychophysics and neurophysiology and allows for a more integrated approach than was previously possible. Offering an accessible account of visual processes and a general survey of ideas about sight, *Vision* serves as a model for problem solving and imaginative inquiry in all areas of scientific endeavor.

David Marr, late of the Massachusetts Institute of Technology, was the author of many seminal articles on visual information processing and artificial intelligence.

VISION



MARR

VISION

FREEMAN

DAVID MARR