Visual Object Representation: Interpreting Neurophysiological Data within a Computational Framework

David C. Plaut
School of Computer Science
Carnegie Mellon University

Martha J. Farah
Department of Psychology
Carnegie Mellon University

Abstract

■ Significant progress has been made in understanding vision by combining computational and neuroscientific constraints. However, for the most part these integrative approaches have been limited to low-level visual processing. Recent advances in our understanding of high-level vision in the two separate disciplines warrant an attempt to relate and integrate these results to extend our understanding of vision through object representation and recognition. This paper is an attempt to contribute to this goal, by using a computational framework arising out of computer vision research to organize and interpret human and primate neurophysiology and neuropsychology. ■

INTRODUCTION

Vision can be characterized as the process of deriving the identities and spatial dispositions of objects in the surrounding environment from the information contained implicitly in retinal images. David Marr (1982) emphasized that understanding a complex information-processing task such as vision requires developing explanations at three levels of analysis: (1) *computational theory*: the purpose of the computation and its justification for the task; (2) *representation and algorithm*: the way that input and output are coded and the algorithm for transforming one into the other; and (3) *hardware implementation*: the way that the representations and algorithm are physically realized. Each of the many disciplines engaged in vision research can be characterized in terms of the primary levels of analysis in which its explanations are couched. Computer science focuses on the first and second levels. Its goals are to characterize the nature of the information that is available to visual processes and the constraints on these processes that arise from the environment and the demands of vision, as well as to develop representations and algorithms that efficiently carry out these computations. Neuroscience works primarily, but not exclusively, at the third level, studying how neural structures physically implement the processing of visual information.

The fruitfulness of sharing constraints and ideas across levels of analysis has been most convincingly demonstrated in the study of low-level vision. Substantial insights about the processes that extract color, edge, motion, and spatial frequency information from retinal images have come from combining computational and neurophysiological constraints. For example, a solution to the computational requirements that one encounters in preparing an image for subsequent edge-detecting operations in early vision—the necessity for smoothing the image and finding intensity gradients at different scales—was suggested by physiological studies of retinal ganglion cells (Marr & Hildreth, 1980). Reciprocally, the interpretation of the function of these cells in terms of a precise computational theory of edge detection has generated further implications for their physiology, which have in turn guided physiological research (Poggio, 1983; Richter & Ullman, 1986).

When we turn to higher levels of visual processing, such as those concerned with object recognition, we see much less interplay between the different levels of analysis. In large part this is a result of the state of knowledge about object recognition *within* each level of analysis. In computer science, there is a sharp contrast between the fairly general and powerful methods for low-level image processing and the generally more limited, special-purpose systems that have been developed for object rec-

ognition (e.g., Ballard & Brown, 1982; Horn, 1986). In neuroscience, single unit recordings and lesion studies have yielded a detailed, coherent account of many aspects of visual processing from the retina through striate cortex (e.g., Lennie, 1980; Livingstone & Hubel, 1984; Lund, 1988) and for many of the prestriate cortical areas (e.g., Maunsell & Newsome, 1987; Zeki, 1978; Van Essen, 1985). Yet our knowledge of the neural mechanisms that underlie object recognition is in a relatively piecemeal state.

Given that our understanding of object recognition lags behind our understanding of low-level vision in each of these separate disciplines, it is not surprising that interdisciplinary approaches would be delayed. Nevertheless, we believe that enough is currently known about object recognition within each discipline to warrant an attempt at interdisciplinary synthesis. Current computational theories of object recognition can provide a much-needed theoretical framework for interpreting the findings of visual neurophysiology, and, reciprocally, the empirical results of neurophysiology can provide important constraints on computational theories of vision.

In the next section we present a computational framework for the design of object representations. Following that, we review the major results from the neurophysiology of object recognition, including lesion studies in humans and monkeys and single cell recording studies. Finally, we consider interpretations of the physiological data in terms of computational issues in object representation and the implications that these data have for the computations being carried out by the visual system.

COMPUTATIONAL ISSUES IN OBJECT REPRESENTATION

The way in which information is represented can greatly affect how easy it is to do different things with it. For example, multiplication is straightforward when numbers are represented in the Arabic base 10 numeral system, whereas it is quite awkward when they are represented in the Roman numeral system. For a given computation, a good representation produces descriptions that make important information easy to access while making irrelevant or confounding information difficult or impossible to access. The computational criteria for the design of an adequate visual object representation must come out of an understanding of what types of information need to be made explicit, and what types can be made implicit or even discarded, for the purposes of object recognition.

Criteria for Object Representation

Computer vision researchers (Hoffman & Richards, 1985; Marr & Nishihara, 1978) have developed a number of important criteria for object representations: scope, uniqueness, stability, sensitivity, and accessibility. These criteria form the basis of a framework for evaluating proposals of object representations for both machine and human vision.

Scope and Uniqueness

An object representation must be capable of producing an adequate description for each recognizable object. For some applications this may be a restricted class of objects (e.g., polyhedra), but humans can recognize a vast range of different types of objects and so the object representation they employ must have extremely general scope. Not all representations can describe a sufficiently broad range of objects. For example, a representation that employs only planar surfaces would be inadequate to describe objects such as a ball or a tree. Although some representations have sufficient scope in theory (e.g., piecewise planar approximations of curved surfaces), in practice the resulting descriptions fail along other criteria.

Given adequate scope it is important that each object have a unique, canonical description within the representation. Two different objects that are given the same description cannot be distinguished. In addition, if the same object can be given different descriptions on different occasions, the system will be faced with the possibly difficult problem of determining at some point in the recognition process whether two descriptions specify the same object. Thus approaches that use multiple representations to extend their scope must solve the additional problem of determining which represention to use in a given situation so as to ensure the generation of a unique description.

Stability and Sensitivity

The similarity between objects should be reflected in the similarity of their descriptions to ensure robustness in the presence of noise, stability over changes in viewing conditions, and natural generalization to novel objects. The stability of the representation guarantees that the system will be relatively immune to the effects of irrelevant variations in the input.

However, even small differences between objects must be representable if they are significant to the goals of the system. Thus stability cannot be bought at the price of *discarding* the more detailed information about an object. Rather, the stable information that captures the more general properties of an object must be *decoupled* from information that is sensitive to the finer distinctions between objects.

Accessibility

It must be possible to derive the description of an object from information that is available to the recognition process. The limited amount of information present in an

image restricts the types of descriptions that can, in principle, be computed from it. Although it might be possible to extend the class of descriptions that is computable in principle by using top-down knowledge about previously recognized objects, this may make the recognition process computationally intractable. A representation that adequately meets all of the previous criteria, but requires information that is unavailable or requires an unreasonable amount of computation, is useless.

A Design Space for Object Representations

The above criteria specify the desired properties of an object representation. It should be clear that adhering to these criteria is a matter of degree and that designing a representation involves making inherent trade-offs between them. Increasing the range of types of objects that can be represented (scope) tends to make it more difficult to ensure that each individual object has a single description (uniqueness). Making a representation more sensitive to important details also tends to make it sensitive to irrelevant ones (i.e., less stable). And in general, improving the scope, uniqueness, stability, and sensitivity of a representation places increasing computational demands on the system, thereby sacrificing accessibility. Because of these trade-offs, there is no single "best" object representation for all recognition tasks, but rather a space of representations that each has particular strengths and limitations.

In characterizing the space of possible object representations, cognitive psychology texts (e.g., Lindsay & Norman, 1977; Reed, 1982) typically describe three classes of object representation: templates, features, and structural descriptions. Unfortunately, the discussion tends to dismiss the first two alternatives as insufficiently general, while remaining vague about the third. Pinker (1985) provides an excellent analysis of the strengths and limitations of object representations based on templates and features (as well as a variation of the template approach based on Fourier analysis) and lays out a set of open issues in the design of an adequate representation based on structural descriptions. In particular, many of the important differences between current theories of object representation can be captured by their positions on three fundamental and roughly independent issues: (1) the nature of the shape primitives used to describe the parts of the object, (2) the spatial reference frame with respect to which the object and its parts are described, and (3) the organization imposed on the components of the object description. In fact, the simple representation schemes mentioned above can be thought of as degenerate cases along some of these dimensions. Template models use a retinotopic reference frame but do not divide the object up into primitives, while feature-based models use well-defined primitives but do not explicitly represent spatial relationships relative to a reference frame.

In the rest of this section, we use the computational criteria presented earlier to characterize the design space for object representations induced by these three issues. The discussion is not intended to be a comprehensive review of object representation in computer vision; rather, it attempts to illustrate the implications and trade-offs involved in the various alternatives that have been proposed to address each of these issues.

Shape Primitives

Much of the power of a representation based on part decomposition derives from decoupling the description of the shapes of parts of an object from the description of how those parts are related spatially. Typically, the shape of each part is described in terms of a parameterized class of *shape primitives*. The scope of a representation depends in large part on the extent to which these primitives are capable of adequately expressing the shape of the parts of objects. In addition, the primitives must be derivable from the image (accessibility) and allow an object to be recognized under different viewing conditions (stability).

There are three basic types of shape primitives used by computer vision systems: contour-based, surface-based, and volumetric. Contour-based primitives include (1) wire-frame models, which represent the significant edges of an object (e.g., Roberts, 1965), (2) skeleton models, which represent the the axes of the major parts of an object (e.g., Blum, 1973), (3) junction models, which represent the arrangement of vertices of a polyhedral object (e.g., Waltz, 1975), and (4) curvature extrema models, which represent the alternation of curvature extrema along significant contours of an object (e.g., Richards & Hoffman, 1985). One of the virtues of contour-based primitives is that they are significantly more accessible than higher order primitives such as surfaces or volumes. For this reason, the majority of existing recognition systems rely heavily on contour-based representations (Ikeuchi, 1987; Lowe, 1987; Huttenlocher, 1988) Unfortunately these representations tend to have limited scope. For instance, Richards and Hoffman's "codons" (see Fig. 1) are perhaps the most general existing contour-based primitives, but they are
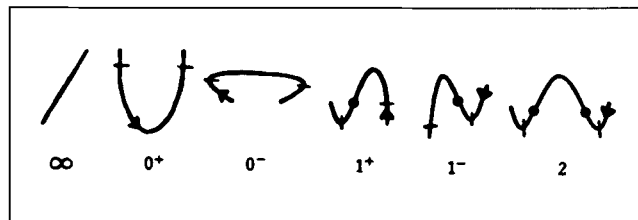

Figure 1. Richards and Hoffman's (1985) primitive codon types, used to describe the occluding contours of an object. Zeroes of curvature are indicated by dots, minima by slashes.

adequate only for objects that can be distinguished solely on the basis of their silhouette.

In contrast, surface-based and volumetric primitives can describe the shape of an object with arbitrary precision. The choice between them is a trade-off between stability and accessibility. A typical surface-based representation consists of local descriptions of the surface properties for small patches of all of the visible portions of an object (see Fig. 2). Representations using volumetric primitives assign three-dimensional descriptions, parameterized for size, shape, and orientation, to each of the major parts of an object. These parts are typically individuated on the basis of elongation or curvature extrema. Examples of volumetric primitives include polyhedra (e.g., Waltz, 1975), spheres (e.g., Badler & Bajcsy, 1978), generalized cylinders (e.g., Nevatia & Binford, 1977), and superquadrics (e.g., Pentland, 1986) (see Fig. 3). Since local surface properties are more directly computable from images than are three-dimensional spatial properties, surface-based primitives place less of a burden on lower level visual processes than do volumetric primitives. On the other hand, the spatial information that volumetric primitives make explicit is much more useful for object recognition than simple surface properties, as it will be more stable under changes in viewpoint. However, deriving volumetric primitives is computationally intensive and relatively few existing computer vision systems employ them (e.g., Brooks, 1981).
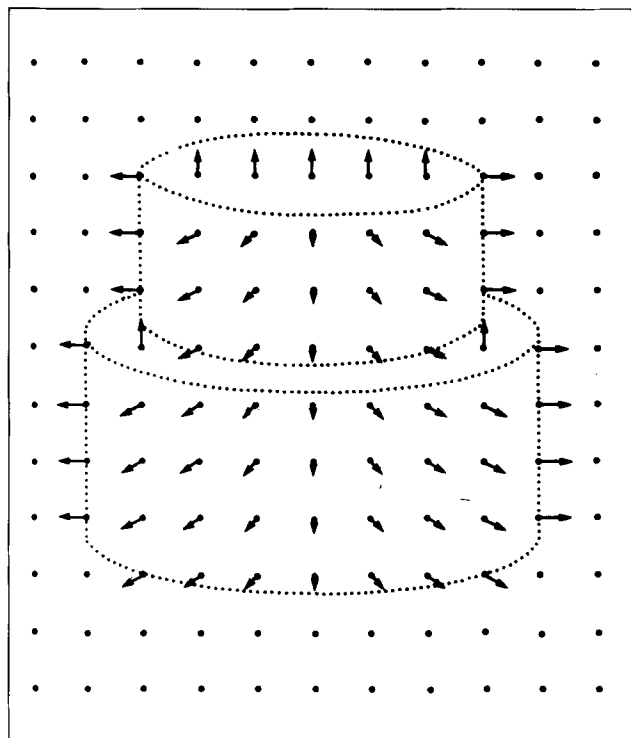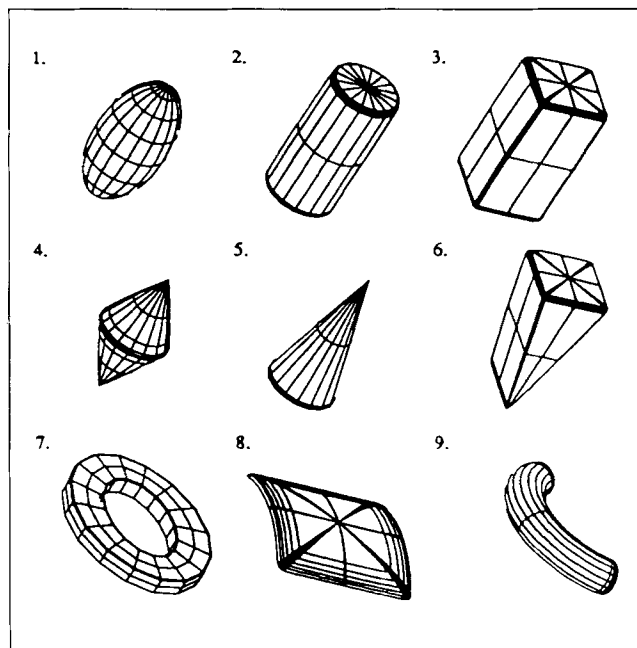

Figure 3. Examples of superquadric volumetric shape primitives. (From Bajcsy & Solina, 1987).

Reference Frame

Regardless of what class of shape primitives is used to describe the parts of objects, the spatial characteristics of these primitives cannot be specified in absolute terms but only with respect to some coordinate system or frame of reference. Hence the choice of spatial reference frame is a fundamental aspect of any theory of object representation.

The initial description of a visual stimulus is represented relative to a frame of reference that is tied to a particular viewpoint; that is to say, it is *viewer-centered*. When the viewpoint changes, either due to an eye movement or change in head and body position, the contents of the representation change. If each object model is also represented in a viewer-centered reference frame, the matching process will be relatively straightforward. Unfortunately, movement of either the object or the viewpoint brings about a change in the derived description so that it will no longer match the same object model, causing the representation to have poor stability.

To achieve stability over changes in the position, orientation, and size of the object with respect to the viewer, a representation should separate the spatial information that is intrinsic to the object (i.e., its shape) from the aspects of the derived description that are idiosyncratic to the current viewpoint. One way to do this is to describe each object model relative to a reference frame that is centered and aligned with itself (i.e., an *object-centered* reference frame, see Fig. 4) rather than to one that can change relative to the object. Recognizing an object involves redescribing the viewer-centered input description relative to the appropriate object-centered
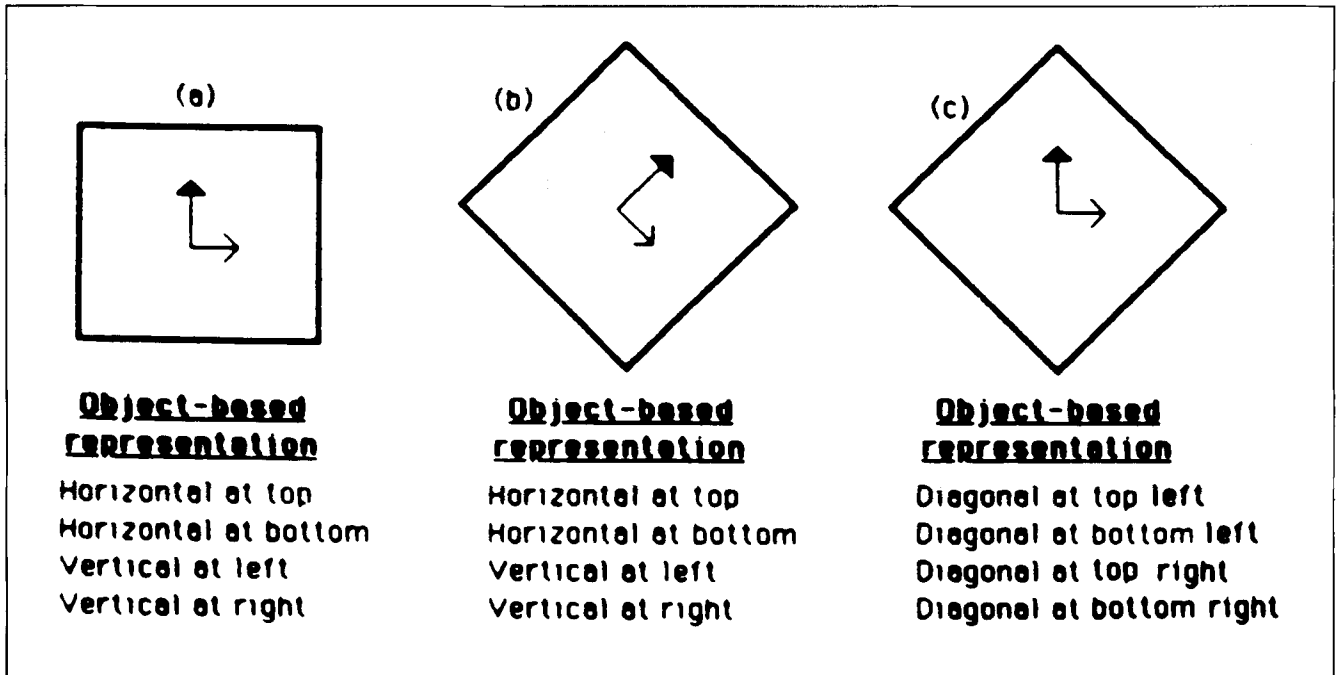

Figure 2. An example of a surface-based representation: the 2½-D sketch of Marr and Nishihara (1978).

Figure 4. An illustration of the use of object-centered frames in shape description. (From Hinton & Parsons, 1988).

reference frame before matching it against object models. Object-centered representations are much more stable than viewer-centered ones because changes in the relation between the object-centered frame and the viewer-centered frame compensate for changes in the relation between the object and the viewer. However, this increased stability comes at the cost of reduced accessibility. The correct object-centered reference frame must be determined *without knowing the identity of the object.* A major current area of research in computer vision is the efficient derivation of object-centered reference frames using viewer-centered properties such as elongation and symmetry (Kanade, 1987; Marr, 1977).

It is important to point out that a reference frame need not be entirely viewer-centered or object-centered. A full three-dimensional reference frame is specified by seven independent parameters: three for its position in three-dimensional space, three for its orientation along three orthogonal axes, and one for scale. Some of these degrees of freedom may be specified relative to the viewer while others are specified relative to the object. For example, objects can be represented by a collection of "characteristic views" (Koenderink & van Doorn, 1979) or "aspect groups" (Ikeuchi, 1987) in which topologically equivalent views of an object (i.e., those with the same set of visible surfaces) are grouped together and given the same representation (see Fig. 5). Because topology changes only with rotation in depth, such a representation can be thought of as involving the assignment of a reference frame that is object-centered in position, scale, and image-plane orientation, but viewer-centered in the two depth orientations. Adopting this type of represen-
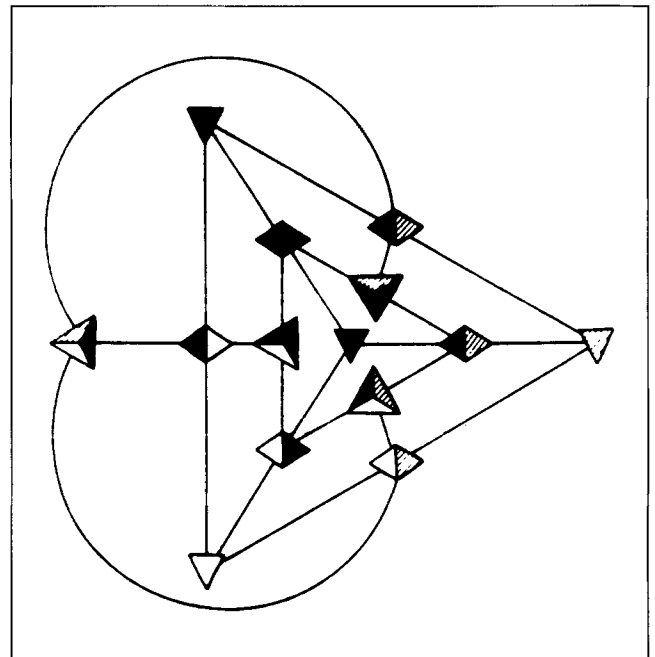

Figure 5. The "visual potential" of a tetrahedron, showing the relationships between its various characteristic views. (From Koenderink & van Doorn, 1979).

tation has certain computational implications. If a large number of views are stored, the computation involved in the matching process increases proportionally. If only a few are stored, then the derived description of the object viewed from some other viewpoint will fail to match any of the stored view and hence go unrecognized.

Thus this type of representation trades off accessibility against scope and stability.

Organization

Given choices for the shape primitives and reference frame used by a representation, the decision on how shape information is to be organized by the representation is still open. The simplest choice is to impose no organization on the information. An example of such a representation is spatial occupancy grids, in which the shape of an object is explicitly represented by a large, undifferentiated collection of volume elements, or "voxels" (Ballard & Brown, 1982). Unfortunately, any imprecision in the lower level processes that derive the voxels will produce significant changes in the resulting object descriptions, making these representations unstable.

Another representation with minimal organization is the use of a separate set of viewpoint-specific templates (e.g., Tarr & Pinker, 1989) for each familiar object. The templates corresponding to a particular object must be grouped together to enable the matched object to be identified, but no organization is imposed within the group.

An alternative way for a representation to organize information is to group information into separate modules, and to explicitly relate these modules to each other (see Fig. 6). Given that objects have visual detail at every spatial scale, and that the parts of objects can often be viewed as objects themselves, the most natural way to organize the modules is hierarchically (e.g., Marr & Nishihara, 1978; Palmer, 1977). The most effective hierar-chical decomposition of an object is in terms of the identities of parts of the object and their spatial relations. The representation of each part consists of (1) its relation to the whole object and to the other parts, and (2) its own hierarchical decomposition, consisting of subparts and their spatial relations. Because objects tend to be larger than their parts, the hierarchy allows information at different spatial scales to be related in a structured fashion. This results in a more stable and sensitive representation because, by grouping together primitives of approximately equal stability (i.e., similar size), the stability of modules using relatively large primitives does not destroy the sensitivity of those using smaller primitives. Hierarchical object descriptions also allow visual processes, such as attention, to naturally vary the spatial scale at which they are directed. However, hierarchical descriptions are more difficult to derive from an image than descriptions based on representations using less structured organizations. In general, greater amounts of organization allow for greater representational stability, but at the cost of accessibility.

Implementational Issues

Thus far, we have shown how computational criteria based on the purposes of object recognition (at Marr's computational level) can constrain the design of an adequate object representation (at the representation and algorithm level) in terms of what shape primitives, reference frame, and organization it uses. Computational vision systems can also be distinguished on the basis of how these representations and their associated algo-
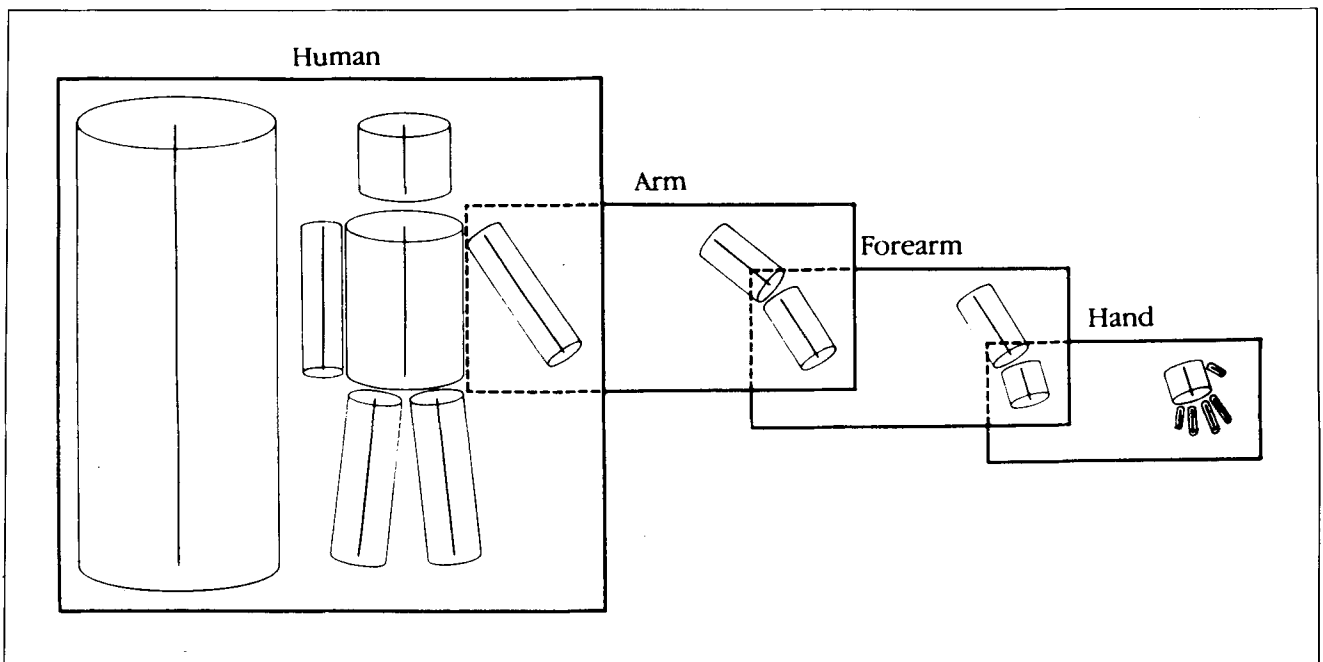

Figure 6. An illustration of hierarchically organized object models. (From Marr & Nishihara, 1978).

rithms are physically implemented in hardware. Although Marr emphasized that the same algorithm can be implemented in quite different technologies, he acknowledged that, among computationally equivalent algorithms, some may be better suited for a particular physical substrate than others.

Approaches to computer vision differ in the type of computational architecture used to implement their representations and algorithms. One class of systems is typified by conventional "symbol manipulation" architectures, in which computation involves the composition of symbol structures by a central interpreter following a stored sequence of program instructions (Newell, 1980; Pylyshyn, 1984). Recently, an alternative computational architecture, known variously as "connectionist models," "neural networks," or "parallel distributed processing," has received considerable attention in cognitive science in general, and computational vision in particular. Computation in these systems takes the form of cooperative and competitive interactions among a very large number of simple, neuron-like computing units (Feldman & Ballard, 1982; Hinton & Anderson, 1981; McClelland, Rumelhart, & PDP Research Group, 1986; Rumelhart, McClelland, & PDP Research Group, 1986). Typically, each unit has associated with it a positive real-valued *state* that loosely corresponds to neural firing frequency. Positive or negative real-valued *weights* on connections between units (corresponding to synapses) determine how the state of each unit influences the states of other units. If the units represent hypotheses about aspects of potential interpretations of the input, the weights on connections between units can encode constraints between these different hypotheses. In this way, the analogue of a "program"—the knowledge about how to process a given input—is not isolated within a central interpreter but rather is encoded throughout the network in the entire set of connection weights. This lack of a separation of program and data is a fundamental difference between connectionist and symbolic architectures (Derthick & Plaut, 1986).

In general terms, computation in connectionist networks occurs in the following way. Initially, input to the system sets the states of some of the units. Then as each unit locally updates its state based on the states of the units with which it is connected, the network as a whole gradually settles into a stable configuration of unit states that represents the interpretation that maximally satisfies the constraints represented by the connection weights given the constraints imposed by the input (Ballard, Hinton, & Sejnowski, 1983; Hinton & Sejnowski, 1983; Hopfield, 1982). Although these networks are poor approximations of actual neurobiology, they may capture many of the important *computational* properties of biological neural networks (Churchland & Sejnowski, 1988).

In this type of computational system, alternative interpretations (e.g., different object identities) are represented as alternative patterns of activity over the *same* set of units. That is, each object activates a number of different units, and each unit participates in representing a number of different objects. This style of "distributed representation" has a number of interesting and useful general properties (Hinton, McClelland, & Rumelhart, 1986). Since there are 2^n possible activity patterns over n units, many more objects can be represented than if each object were represented by a single unit (or separate group of units). Also, similar objects have similar (highly overlapping) representations, so they can have similar effects on other parts of the system in a straightforward way. Furthermore, an unfamiliar object will be represented (i.e., will activate a set of units) in a way that is most consistent with the similarity of its visual appearance to the appearances of known objects. Hence the network generalizes naturally to novel input and can learn to recognize a new object simply by adjusting the weights among units representing similar objects so that the pattern of activity representing the new object becomes stable. Finally, distributed representations are quite resistant to the effects of noise or damage (Wood, 1978; Hinton & Sejnowski, 1986; Hinton & Shallice, 1990; Patterson, Seidenberg, & McClelland, 1990).

Although the implementational characteristics of connectionist networks map naturally onto some aspects of human object recognition, it is important to realize that these advantages in no way eliminate the need to understand and solve the difficult problems at the algorithmic and computational levels.

Summary

To summarize this section, theories of object recognition vary according to their choice of shape primitives, spatial reference frame, and the organization imposed on part representations. These choices can be thought of as defining a space of the possible object representations underlying visual object recognition. Each position in this space involves trade-offs between satisfying the various computational criteria for an adequate object representation that were discussed in the previous section. In general, as representations improve their scope, stability, and sensitivity they sacrifice accessibility, placing greater and greater computational demands on the recognition system. The existence of these trade-offs makes it difficult to choose one type of model as the "correct" model based on computational considerations alone. In addition, computational systems differ in the type of computational architecture they use to implement their representations and algorithms. In the next section we review a set of neurophysiological data that may provide empirical evidence of the design decisions and implementation chosen by the primate visual system. At the same time, the space of alternative models described above will provide a framework for interpreting these data.

NEUROPHYSIOLOGY OF OBJECT RECOGNITION

There are three main sources of evidence about the neural bases of object recognition: clinical studies of brain-damaged humans, lesion studies of animals, and single-cell recording studies of animals. In this section of the paper we will survey the major results obtained with each of these methods.

Clinical Evidence from Brain-Damaged Humans

Damage to the posterior regions of the human brain can result in impairments in visual object recognition. Truly selective deficits in object recognition are known as visual associative agnosias. Patients with associative agnosia are unable to recognize visually presented stimuli despite apparently preserved visual perception and general knowledge of the objects (see Farah, 1990, for a review). For example, they cannot recognize an object by seeing it, but can recognize it readily by touching it or hearing its sound. Furthermore, they can draw an excellent copy of it when it is placed in front of them, which seems to imply that their perception of it is not at fault (see Fig. 7). Associative agnosia is often contrasted with apperceptive agnosia, in which object recognition fails because lower level visual perception is grossly impaired. These patients cannot reliably discriminate a straight line from a curve, or an "X" from an "O." Whatever inferences can be made from such patients about the nature of vision, they will concern relatively early visual processes and not those concerned specifically with object recognition.
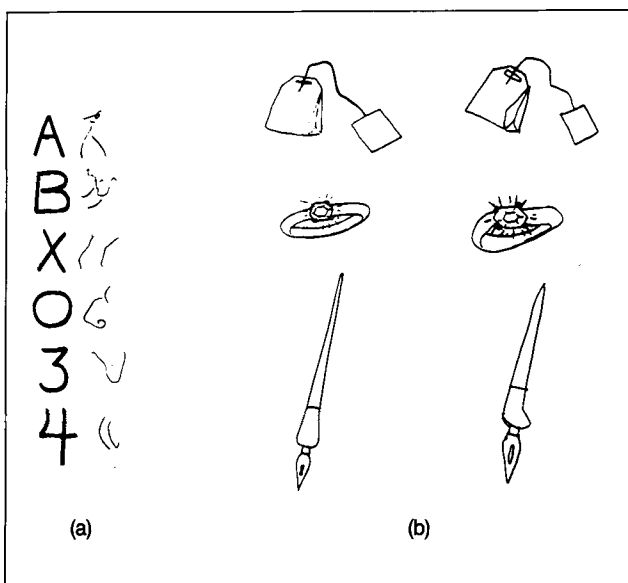
Hence for our purposes we will focus on the characteristics of associative agnosia.

Lissauer (1890, translated in Shallice & Jackson, 1988) originally defined associative agnosia as the inability to access semantic knowledge of objects from truly intact visual representations. Although it is possible that some patients described as associative agnosics do have completely intact perception (see Shallice, 1988, for a discussion of this possibility), in most cases in which the patient's visual capabilities have been systematically studied there is evidence that a subtle visual impairment is responsible for their agnosia. For example, when associative agnosics draw, they do so extremely slowly and laboriously, rendering the copy a line at a time (Humphreys & Riddoch, 1987). Ratcliff and Newcombe (1982) found that their patient M.S. was unable to relate different views of the same object to one another (see Fig. 8). They also noted that M.S. was unable to discriminate between "possible" and "impossible" figures (Gregory, 1970). This task has nothing to do with recognizing previously familiar objects, but merely requires the construction of a visual representation of the structure of a whole object. However, the construction of this visual representation is undoubtedly a prerequisite for recognition. They therefore argue that in their case, at least, associative agnosia results from an inability to construct a coherent structural description of visual stimuli.

Riddoch and Humphreys (1987) reached a similar conclusion with their patient H.J.A. In one study, they presented him with an "object decision task," in which drawings had to be classified as real objects or as made-up objects created by grafting together parts of real objects. H.J.A. was impaired at this task, but he was paradoxically better at performing it when the drawings were filled in and presented as silhouettes. Riddoch and Humphreys interpret this as evidence that his problem is an inability to integrate separate visual features together into

Figure 7. Examples of the copying abilities of (a) apperceptive agnosics (from Benson & Greenberg, 1969), and (b) associative agnosics (from Farah, Hammond, Levine, & Calvanio, 1988). None of the pictures was correctly identified.
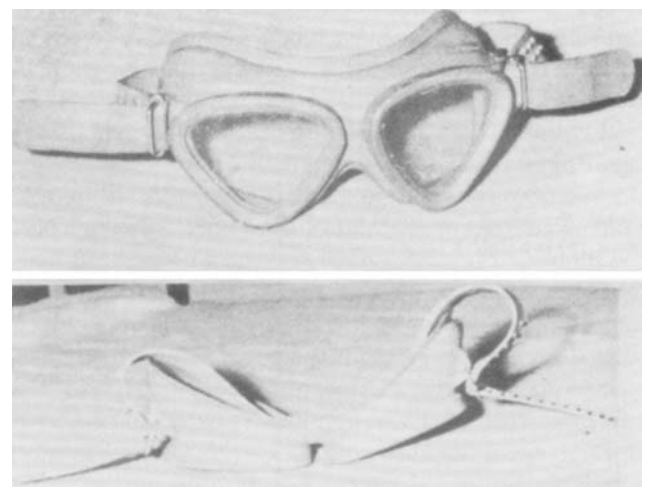

Figure 8. An example of an object photographed from a usual view (top) and an unusual view (bottom). (From Warrington, 1982).

a visual representation of the whole. The greater number of details in the drawings, compared to the silhouettes, made this a harder task for him. Humphreys and Riddoch (1987) also showed that H.J.A. is impaired in visual search experiments. In experimental contexts in which normal subjects can benefit from the good configuration of the stimulus array, H.J.A. shows the same slow, serial search as when the stimulus locations are random.

Levine and Calvanio (1989) report the results of a series of standardized, factor-analyzed visual/spatial tests with an agnosic patient, L.H. They found that he was impaired mainly on the tests that emphasize the "visual closure" factor. These tests require synthesizing fragmented or partially occluded stimuli into a "whole" (see Fig. 9). In contrast, L.H. performed *better* than normals on tasks that emphasize the "flexibility of closure" factor, in which subjects must find shapes hidden within larger patterns. Normal subjects find this task difficult because the hidden shape often does not correspond to a "good" or natural part of the larger whole. These results are consistent with the idea that L.H., like M.S. and H.J.A., does not automatically see objects as coherent wholes. In sum, the available data from three studies of the visual capabilities of associative agnosics all point to an impairment in their ability to see the overall structure of an object, and the relation of its parts to its overall structure.

Prosopagnosics have a relatively circumscribed recognition impairment that mainly affects the recognition of faces. They may be able to read, recognize most common objects, photographs, and drawings, but be so profoundly impaired at face recognition that they cannot recognize their own family by sight, or even themselves in a mirror. Like associative agnosics, prosopagnosics have traditionally been described as having normal vision, but evidence is now accumulating to the contrary. For example, some prosopagnosics perform within normal limits on the Benton & Van Allen test of facial discrimination, in which unfamiliar faces must be matched across changes in perspective and lighting (Benton & Van Allen, 1972). However, when the time required to perform the test, and the manner of performing the test,

are taken into account, the "normalcy" of these patients appears questionable. Typically, they resort to slow, serial checking of the faces, verifying one feature at a time (Ellis & Young, 1987). Again, this is broadly consistent with an impairment in seeing how the individual parts of an object relate to the whole.

Although the neuropsychological studies summarized above suggest that an impairment in object representation underlies associative agnosia, their usefulness is limited in a number of ways. First, with few exceptions, research with these cases has been largely descriptive. Although all of the studies seem to indicate a difficulty in representing the overall shape of a complex object, they do not allow precise inferences regarding the nature of the underlying functional impairment. Second, the appropriate cases are quite rare, and the exact locations of their brain damage is variable and often unknown. Although bilateral inferior temporal-occipital damage is common (Alexander & Albert, 1983), some authors have described agnosia-like syndromes following unilateral temporal-occipital lesions and lesions affecting predominantly parietal areas (Warrington, 1982). Experimental work with animals, summarized in the following section, allows greater control over lesion localization and has generally included more systematic investigations of the functional nature of the deficit.

Lesion Studies in Animals

The earliest experimental work on the neurophysiology of object recognition in animals involved the bilateral surgical ablation of different parts of the occipital and temporal lobes of monkeys. Kluver and Bucy (1937, 1939) discovered that the complete removal of both temporal lobes causes a rather complex disruption of monkeys' social, sexual, and eating behavior, known as the "Kulver–Bucy syndrome," of which a failure to recognize visual stimuli is one aspect. In the decades that followed, researchers attempted to fractionate this syndrome and to narrow down the particular areas of the temporal lobe involved in visual abilities (Blum, Chow, & Pribram, 1950;

Figure 9. Examples of stimuli in a recognition task that stresses the "visual closure" factor. (a) is a flag; (b) is a hammer head. (From Elkstrom, French, & Harman, 1976).

(a)   (b)
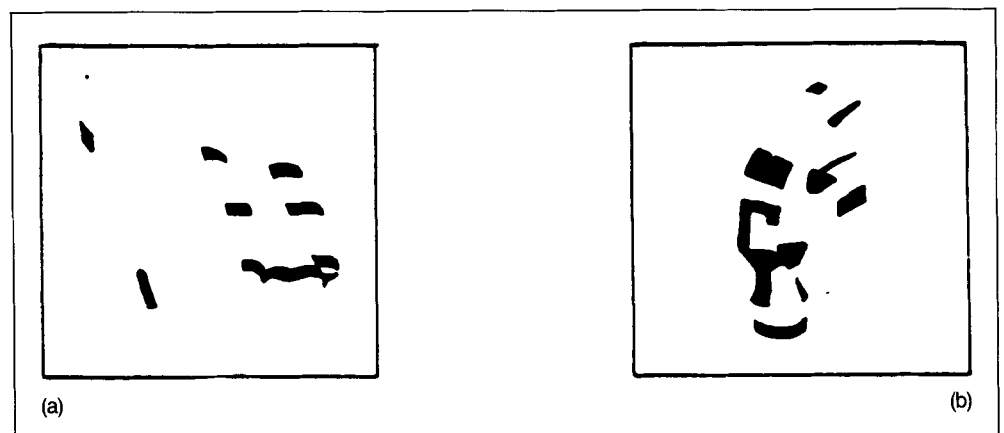
Chow, 1951, 1952). It was eventually determined that lesions confined to the neocortex of the inferior temporal gyrus (inferotemporal cortex, or IT see Fig. 10), corresponding roughly to area TE of von Bonin and Bailey (1947), are sufficient to produce visual deficits (Mishkin, 1954, 1966; Mishkin & Pribram, 1954). A great deal of subsequent research has been aimed at precisely

Figure 10. The location of cortical visual areas in the macaque, including posterior and anterior inferotemporal areas (PIT and AIT), viewed (a) laterally, (b) medially, and (c) with the superior temporal sulcus opened. (From Maunsell & Newsome, 1987.)

characterizing nature of the visual impairment produced by IT lesions in monkeys. Most of these investigations did not test visual object recognition per se (as is done with human visual agnosics), but rather tested the ability of IT-lesioned monkeys to *learn to discriminate* among visual stimuli (see Levine, 1982, for a detailed comparison of the two testing conditions). To explicitly relate these results to object recognition it will help to review the type of task typically used in lesion studies.

In visual discrimination experiments, the monkey is rewarded for responding differentially (e.g., by button press) to a particular visual pattern, which is presented with one or a number of distracting stimuli that are often visually similar to the rewarded pattern. In one type of visual discrimination task, the "simultaneous forced-choice task," the rewarded pattern and distractors are presented concurrently. In another common version, the "delayed match-to-sample task," the rewarded pattern is presented and then removed, and then it is presented again along with distracting stimuli. In this way which pattern is rewarded can be varied from trial to trial. In a lesion experiment, the experimental, normal, and operated control groups may be compared in terms of the number and type of errors made, the number of trials required to learn the discrimination to some performance criterion, or the extent to which the animals in the group were able to perform the task at all.

In experiments involving bilateral lesions of IT in monkeys, the most striking result is a severe impairment in learning visual discriminations in tasks such as those described above. These monkeys require many more learning trials to reach criterion than normal or operated control monkeys (e.g. Blum et al., 1950; Mishkin, 1966; Pribram, 1954). Although the visual discrimination deficit is generally demonstrated in the context of tasks that require new learning, IT-lesioned monkeys also show a severe loss in retention of a discrimination learned pre-operatively (e.g., Dean & Weiskrantz, 1977; Gross, 1978; Pribram, 1954), an impairment more closely analogous to human visual object agnosia.

Monkeys with IT lesions do not simply learn a normal discrimination more slowly; they appear to use stimulus features abnormally. Butter, Mishkin, and Rosvold (1965; Butter, 1968) found that after IT-lesioned monkeys had learned to discriminate a grating of a particular orientation and color from other patterns they were more likely than normals to inappropriately respond to stimuli of a similar orientation or color as the original rewarded stimulus. Analogous results obtain for the generalization of discriminations involving angles (Blake, Jarvis, & Mishkin, 1977). Iwai (1985) presented a series of experiments demonstrating that IT-lesioned monkeys learn visual discriminations by relying on idiosyncratic lower level aspects of the stimuli. For example, in discriminating between a triangle and a circle, IT-lesioned monkeys learned to respond to the fact that the bottom line of the triangle was parallel to the bottom edge of the back-
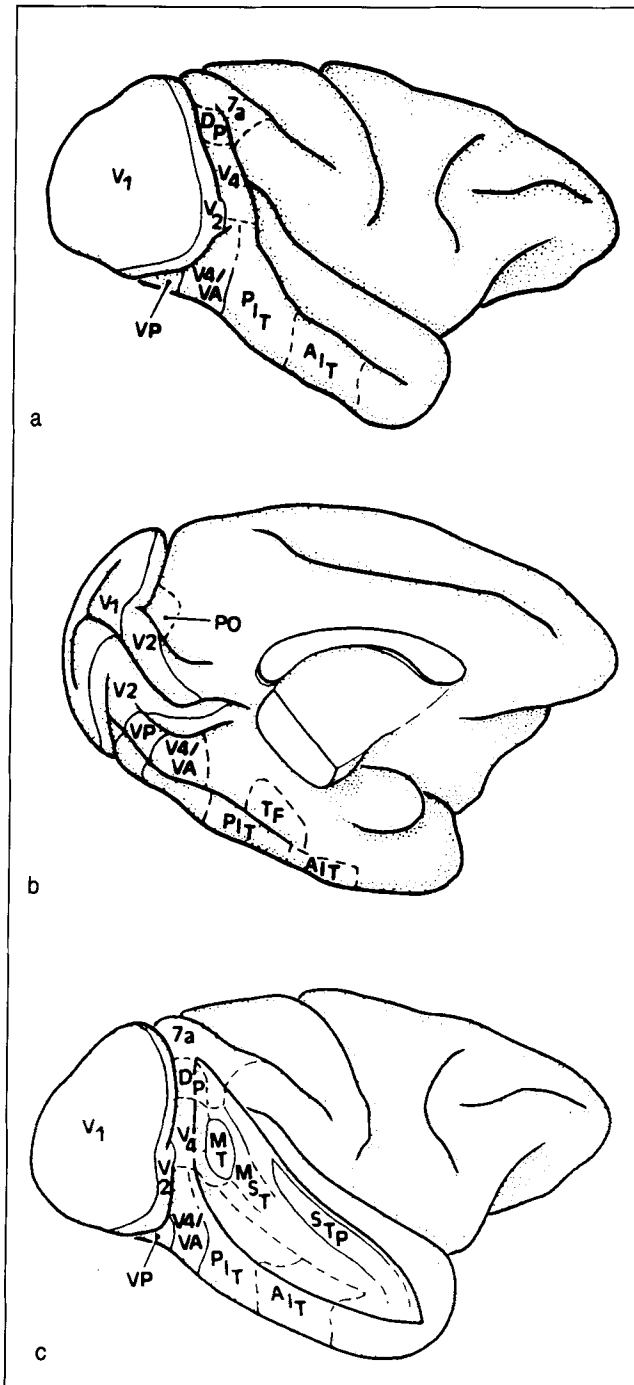
ground plaque; when this relationship was eliminated (by rotating the plaque relative to the triangle but leaving the patterns unchanged) IT-lesioned monkeys, but not normals, lost the discrimination. Gaffan, Harrison, and Gaffan (1986a) found that following IT lesions, the performance of monkeys trained on a serial reversal learning task (in which the reward association of two stimuli are varied) recovers to preoperative levels, in contrast to those trained on a more conventional discrimination learning task (involving a new pair of stimuli for each problem). They suggest that IT lesions reduce the number of attributes that are used to describe stimuli, so that tasks involving only a few stimuli are relatively unimpaired.

Lower level sensory deficits such as field defects, acuity losses, or raised visual thresholds have not been found following IT lesions (Cowey & Weiskrantz, 1967; Mishkin & Weiskrantz, 1959; Weiskrantz & Cowey, 1963). Furthermore, animals with such defects (generally as a result of striate lesions) are less impaired than IT-lesioned monkeys on many visual discrimination tasks (Butter et al., 1965; Wilson & Mishkin, 1959). Hence the role of IT in visual discrimination learning does not directly involve these low-level image properties. It has also been shown that the deficit is exclusively visual: olfactory discrimination (Brown, 1963), tactile discrimination (Wilson, 1957), and auditory discrimination (Weiskrantz & Mishkin, 1958) remain unimpaired. In contrast, lesions of higher cortical areas that receive input from IT (e.g., the temporal pole and superior temporal sulcus) either produce no visual discrimination deficit, or produce deficits in multiple modalities rather than in vision alone (Brown, 1963; Mishkin, 1972). This implies that IT is concerned exclusively with the processing of visual stimuli, and that it is the final processing station in the brain for visual stimuli within the visual system proper.

IT receives most of its input from a particular part of prestriate cortex, roughly corresponding to area TEO of von Bonin and Bailey (Kuypers, Szwarcbart, Mishkin, & Rosvold, 1965). This area has been called "foveal prestriate" cortex (Cowey & Gross, 1970) because of its disproportionate representation of foveal visual stimuli. Predictably, lesions of foveal prestriate cortex also impair performance on visual discrimination learning tasks (Cowey & Gross, 1970; Heywood & Cowey, 1987; Iwai & Mishkin, 1968, 1969). However, the character of these visual discrimination deficits is quite different from those caused by IT lesions. In general, monkeys with foveal prestriate lesions are more severely impaired than those with IT lesions. They fail on all but the simplest discriminations (Iwai & Mishkin, 1968), and have worse postoperative retention of a learned discrimination than do IT-lesioned monkeys (Cowey & Gross, 1970; Iwai & Mishkin, 1969). However, while foveal prestriate lesioned monkeys are worse at learning to make difficult discriminations, IT-lesioned monkeys are worse at "concurrent

discrimination" learning, in which a number of simple discriminations, which the monkey would have no trouble learning separately, are interleaved and must be learned in parallel (Cowey & Gross, 1970; Iwai & Mishkin, 1968; Mishkin, 1972). In general, monkeys with IT lesions are more distracted by intervening tasks (Dean & Cowey, 1977; Gross, Cowey, & Manning, 1971; Iversen, 1970) while those with foveal prestriate lesions are more distracted by the removal of redundancy (Wilson & Kaufman, 1969) or addition of irrelevant features to a stimulus (Dean & Cowey, 1977; Gross et al., 1971; Iwai & Mishkin, 1969). Levine (1982) and Heywood and Cowey (1987) suggest that the pattern of deficits following foveal prestriate lesions in monkeys is analogous to apperceptive agnosia in humans.

Initial attempts at interpreting these results characterized the different functions of foveal prestriate cortex and IT, respectively, as "discrimination vs. visual memory" (Iwai & Mishkin, 1968; Mishkin, 1972), "identification vs. encoding" (Wilson, Kaufman, Zieler, & Leib, 1972), and "perceptual vs. associative" (Cowey & Gross, 1970; Gross, 1973). Although these simple dichotomies served to organize thinking and guide further research, additional experimentation made it clear that they were inadequate explanations. For example, the hypothesis that IT subserves visual memory was challenged by a series of delayed match-to-sample tasks in which the time between the initial presentation of the rewarded stimulus and its later presentation among distractors was varied (Dean, 1974). The demand on visual memory increased with longer delays, and normal monkeys committed progressively more errors. Yet monkeys with IT lesions who, after extensive training, learned the task at zero delay were no more severely affected by increasing delays than normal monkeys. This suggests that the function of IT is not visual memory per se (Gaffan et al., 1986a).

Most recent lesion studies have focused on attempting to determine the types of information that are and are not represented in IT by varying the relationship between the rewarded stimulus and distracting stimuli in visual discrimination tasks. To the extent that stimuli that differ along a particular visual dimension are less discriminable to monkeys with IT lesions, this visual dimension is presumably represented in IT. Conversely, to the extent that stimuli are equally discriminable to monkeys with and without IT lesions, the dimension of difference is arguably not represented in IT.

Position

The *retinal position* of a stimulus appears to be irrelevant to IT representations (Gross & Mishkin, 1977; Seacord, Gross, & Mishkin, 1979). Monkeys with bilateral, but not unilateral, IT lesions show impaired interhemispheric transfer of a learned visual discrimination (i.e., impaired

generalization across the two hemifields). This implies that IT is necessary for stimulus equivalence between the two visual hemifields and, presumably also, for equivalence between retinal positions within a hemifield.

Size

The *size* of stimuli is another property that appears to be represented abnormally in IT-lesioned monkeys. Humphrey and Weiskrantz (1969) trained monkeys to discriminate two disks at varying distances on the basis of their physical size. After IT lesions the monkeys were unable to relearn the task and instead responded on the basis of either retinal size or distance. Ungerleider, Ganz, and Pribram (1969) later replicated and extended these findings on IT lesions and size constancy. Weiskrantz and Saunders (1984) found that after training normal and IT-lesioned monkeys to discriminate a three-dimensional object paired with a large number of distractors, the lesioned monkeys were impaired relative to normals in discriminations involving larger and smaller versions of the rewarded object. These results imply that whereas IT is not required for the representation of retinal size or distance, it is required for size constancy. In contrast, Holmes and Gross (1984b) found that IT-lesioned monkeys showed normal generalization to scaled versions of a stimulus (a block uppercase letter "J") that had to be discriminated from a single fixed distractor (a block Greek letter "π"). However, these results can be explained if we assume, as suggested by the work described above, that the IT-lesioned monkeys were simply relying on lower level cues, such as the curved segment of the "J", that would be present in scaled versions of the letter and are sufficient to distinguish them from the particular distractor used. Under this interpretation these results do not conflict the claim that retinal size information is not represented in IT.

Orientation

Interpreting the results on the representation of stimulus orientation is far less straightforward than for the previous visual dimensions because changing the orientation of a stimulus also tends to change which features of the stimulus are visible or salient. Given the evidence that IT-lesioned monkeys tend to rely more on idiosyncratic stimulus features than do normals, discrimination deficits in these experiments may reflect a difference in sensitivity to the appearance of stimulus features rather than to orientation per se. To tease apart these effects, it is important to distinguish image plane (i.e., frontoparallel) orientation from orientation in depth.

Changes in *image plane orientation* do not affect the visibility of stimulus features but can change their salience, since monkeys tend to pay more attention to the part of the discriminanda closest to the response site

(Meyer, Treichler, & Meyer, 1965). In contrast with the conventional finding that IT lesions impair discrimination between different stimuli (so-called "different pattern" discriminations), Gross (1978) found that monkeys with IT lesions are relatively unimpaired at discriminating between simple two-dimensional patterns, such as digits, that differed only by a rotation of 90 or 180° ("rotated-pattern" discriminations). Holmes and Gross (1984a) replicated these results, but did find that IT-lesioned monkeys were worse than normals at discriminating patterns differing only by rotations of 30 or 45° (see Fig. 11). Holmes and Gross also obtained essentially similar results for three-dimensional objects, such as small colored toys, rotated only in the image plane. Thus monkeys with IT lesions are worse than normals at discriminations involving small, but not large, differences in image plane orientation.
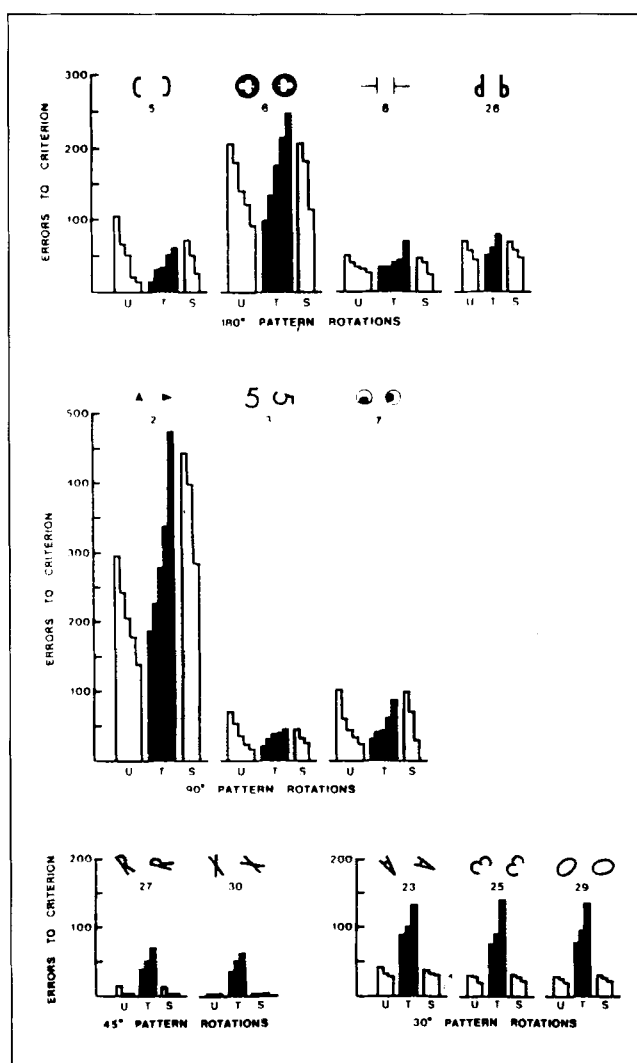

Figure 11. Orientation discrimination performance of monkeys with bilateral IT lesions (T) relative to unoperated controls (U) and those with bilateral lesions of lateral striate cortex (S). (From Holmes & Gross, 1984a.)

Interpreting these results requires separating two effects. First, normal monkeys find rotated-pattern discriminations more difficult than different-pattern discriminations, presumably because they tend to ignore differences in orientation in comparing shapes. Second, the performance of IT-lesioned monkeys at rotated-pattern discriminations improves as the rotation angle is increased (i.e., as the patterns become more discriminable based on low-level features). For rotated-pattern discriminations involving large rotations, IT-lesioned monkeys are at their best and hence are unimpaired *relative* to the "impaired" normals. At smaller rotations, IT-lesioned monkeys have increasing difficulty relying on lower level feature differences, and so their *relative* deficit returns. Under this interpretation, normal monkeys have difficultly responding on the basis of image plane orientation differences, and hence this visual dimension does not appear to be represented in IT.

Orientation in depth might be expected to be treated differently by the visual system from orientation in the image plane, as depth rotations generally change the appearance of a stimulus in more complex ways, revealing previously hidden surfaces and occluding previously visible ones. Indeed, Weiskrantz and Saunders (1984) found that monkeys with IT lesions showed reduced transfer from a learned discrimination to one involving a 90° rotation in depth. A possibly conflicting result comes from the work of Holmes and Gross (1984b), who failed to find a generalization deficit in IT-lesioned monkeys for 60° depth rotations. However, the block letters "J" and "π" used in the Holmes and Gross study would retain their discriminability based on lower level cues under 60° depth rotations. Thus, a preliminary conclusion from the available data would be that IT is required for the representation of equivalence over depth rotations.

Depth

The perception of *depth* per se seems to depend to some degree on IT. Cowey and Porter (1979) demonstrated that IT lesions impair the ability of monkeys to discriminate depth in red-green anaglyph random-dot stereograms when the binocular correspondence is reduced. Holmes and Gross (1984b) found that after learning a discrimination involving three-dimensional stimuli, IT-lesioned monkeys generalize more poorly than normals to discriminations involving a two-dimensional (2D) version of the original rewarded object. This result could be interpreted as implying that IT is involved in perceiving stimuli as three-dimensional (3D) objects rather than as 2D images, if one assumes that the similarity between the 3D and 2D versions of the stimulus will be greatest when they are viewed as representing 3D objects. The fact that discriminations between line orientation in the image plane do not appear to involve IT (Gross, 1978)

is consistent with the hypothesis that IT plays a special role in representing depth.

Illumination

The shadows that an object casts across itself as a function of the location of the *source of illumination* can change the appearance of an object. Whereas normal monkeys do not show any difficulty generalizing across different conditions of illumination, Weiskrantz and Saunders (1984) found that IT-lesioned monkeys were impaired at this generalization (see Fig. 12).

Enantiomorphy

A surprising result concerns the preserved ability of IT-lesioned monkeys to make *enantiomorphy* (mirror-image) judgments. Among the results of Cowey and Gross (1970) and Gross (1973) are examples of pairs of stimuli that normal monkeys find extremely difficult to discriminate, yet on which monkeys with IT lesions are no worse. Each of these stimulus pairs consisted of lateral mirror images. Further experimentation (Gaffan et al., 1986a; Gross, 1978; Gross, Lewis, & Plaisier, 1975) con-
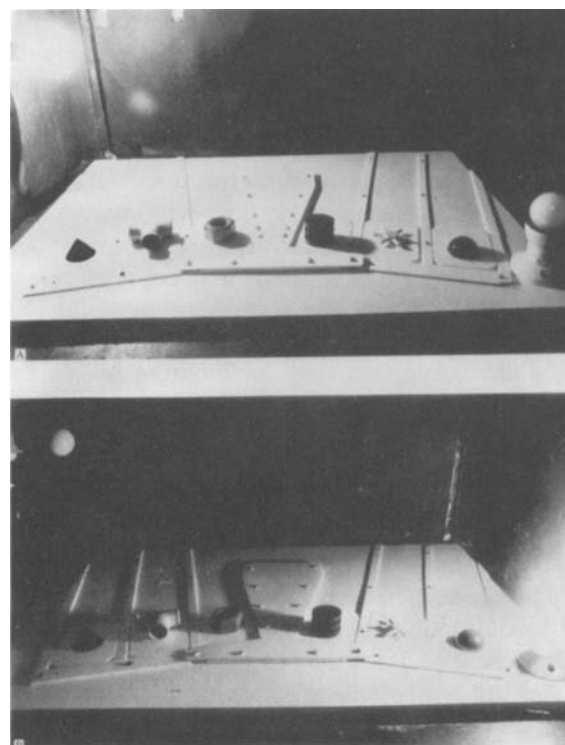

Figure 12. Effects of illumination change on stimuli in a six-alternative forced-choice task used by Weiskrantz and Saunders (1984).

firmed that monkeys with IT lesions are as good as normals at discriminating stimuli that differ only in handedness. (see Fig. 11). These results make sense if the handedness of an object is not explicitly represented in IT; normals find these discriminations unusually difficult because both patterns have the same description in IT, while lesioned monkeys rely on lower level descriptions in which the enantiomorphs are often quite different.

The studies reviewed above show that IT is necessary for representing shape independent of its retinal size, location, handedness, three-dimensionality, and, for the most part, orientation. This characterization of the properties of representations in IT helps to explain the earlier findings that seemed to implicate it in visual learning and memory. Because the shape representations in IT are more highly abstracted from the stimulus array than earlier representations in striate and prestriate cortex, they provide a more "concise" representation of stimulus shape (i.e., leaving out irrelevant information about position, size, etc.). The more concise a representation one has available, the greater the mnemonic capacity for retaining information (Miller, 1956).

Although lesion studies in animals provide information about the nature of the stimulus representations in IT, this information depends on fairly indirect inferences from animals' behavior in complex tasks. Furthermore, it has already been noted that lesioned animals may develop idiosyncratic strategies for performing these tasks. An advantage of single-unit recordings is that one can directly observe the response of the visual system to a variety of stimuli, independent of postvisual cognitive processing required for performing behavioral tasks.

Single-Cell Recording

Early investigations of the electrophysiology of IT recorded from single cells in anesthetized monkeys during the presentation of simple visual stimuli, such as colored oriented bars (Gross et al., 1967, 1969, 1972). The majority of neurons in this area are visually sensitive, with large receptive fields (about 26° in diameter on average), extending into both visual hemifields, and always including the fovea. However, these cells do not seem to be sensitive to the association of the stimulus with reward (Rolls, Judge, & Sanghera, 1977; Sato, Kawamura, & Iwai, 1980). In contrast with earlier visual areas, IT shows no visuotopic organization (Desimone & Gross, 1979), although cells with similar response properties tend to cluster (Fuster & Jervey, 1982). The responses of IT cells are enhanced during discrimination tasks as compared with conditions in which the monkey need attend only to the stimuli (Richmond & Sato, 1987), and become larger and more selective as the difficulty of the discrimination increases (Spitzer, Desimone, & Moran, 1988).

Researchers have had great difficulty determining the optimal stimulus for many IT cells. Although many cells respond well to virtually any stimulus, other cells are selective along a particular visual dimension and relatively insensitive along others. Some cells have been found that appear to respond quite selectively for a particular complex stimulus, such as forceps, a brush, a monkey hand, or a face. Further research has revealed that the superior temporal sulcus (STS) contains a relatively high proportion of cells selective for faces (Baylis, Rolls, & Leonard, 1985; Bruce, Desimone, & Gross, 1981; Desimone, Albright, Gross, & Bruce, 1984; Perrett, Rolls, & Caan, 1979, 1982: Perrett, Smith, Potter, Mistlin, Head, Milner, & Jeeves, 1985; Rolls, 1984; Rolls & Baylis, 1986; Rolls et al., 1977; Yamane, Kaji, & Kawano, 1988) (see Fig. 13). Of the cells in the temporal cortex that respond selectively to complex stimuli, the strongest and most selective responses are to faces (Baylis et al., 1985).

To understand how visual information is represented in IT, much recent work has focused on precisely characterizing the way in which the response properties of visually responsive IT cells in awake, behaving monkeys are (or are not) affected by changes in the stimulus along visual dimensions such as shape, texture, color, size, and orientation. The stimuli used in these studies included simple bars of varying lengths and widths, two-dimensional shapes and patterns, and complex three-dimensional objects. Consistent with the lesion studies, the general conclusion that has emerged is that IT cells are sensitive to aspects of the stimulus that reflect stable physical properties of the object while remaining insensitive to aspects that are specific to the particular viewing conditions. Many IT cells respond selectively along the dimensions of shape, color, and texture (Desimone et al., 1984, 1985; Richmond, Optican, Podell, & Spitzer, 1987; Schwartz, Desimone, Albright, & Gross, 1983) while they are relatively unaffected by changes in contrast (Rolls & Baylis, 1986; Sato et al., 1980), retinal position (Desimone et al., 1984, 1985; Miyashita & Chang, 1988; Schwartz et al., 1983), retinal size (Desimone et al., 1984; 1985; Iwai, 1985; Miyashita & Chang, 1988; Perrett et al., 1982, 1985; Rolls & Baylis, 1986; Sato et al., 1980; Schwartz et al., 1983), and image plane orientation (Desimone et al., 1984; Iwai, 1985; Miyashita & Chang, 1988; Perrett et al., 1985). It should be pointed out that individual cells do not show perfect invariance in their responses over changes along these stimulus dimensions; it is only the population of responses that collectively contains sufficient information to factor out the effect of these variables (Baylis et al., 1985).

Some IT cells do appear selective to the orientation of an object in depth (as opposed to image-plane orientation). Rolls et al. (1977) found cells whose activity varied for different views of an object, while Desimone et al. (1984) and Perrett et al. (1985) found face-selective cells in STS that preferred frontal over profile views, while others had the opposite selectivity.

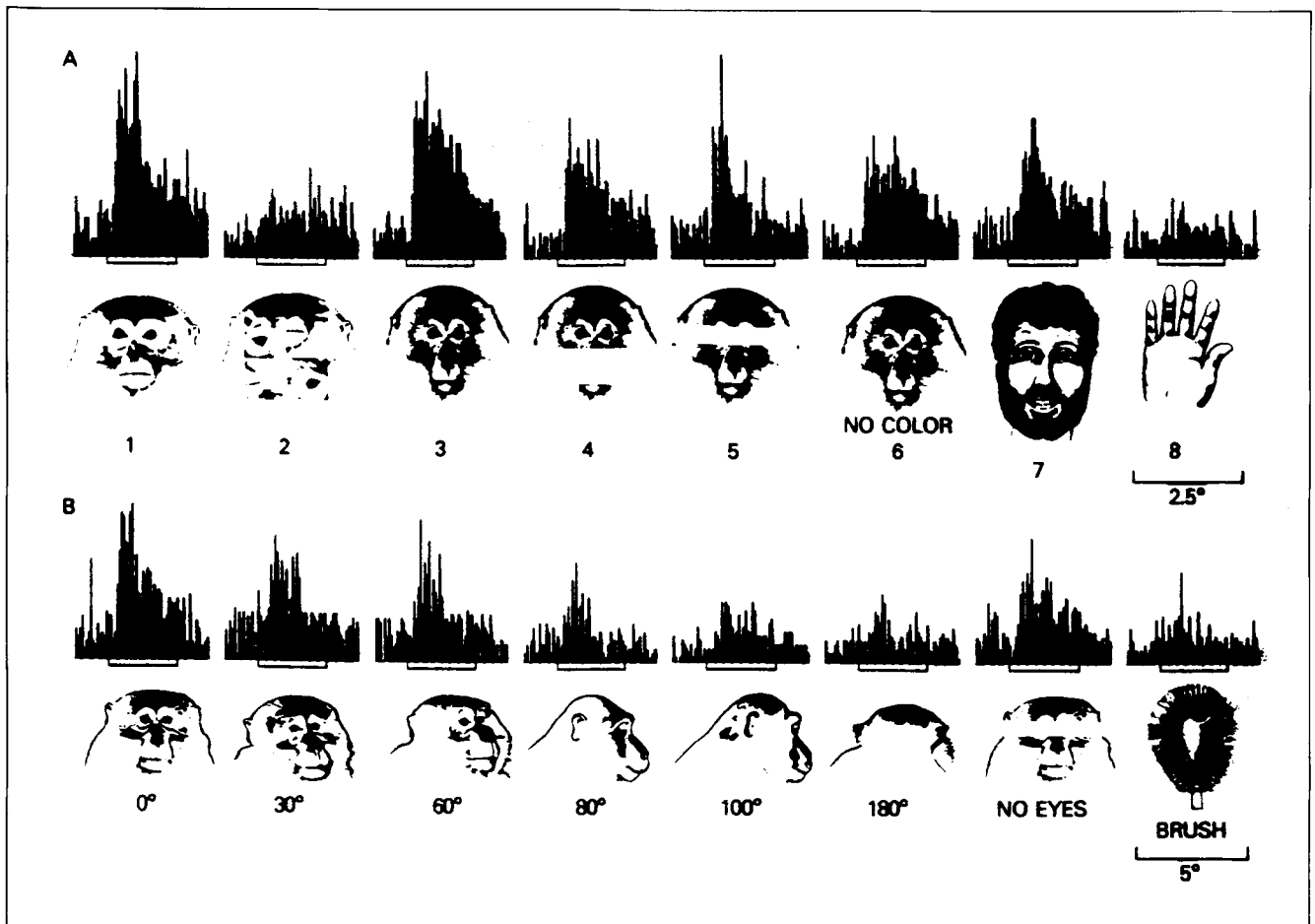An interesting set of results concerns the responses of

Figure 13. Responses of a face-selective cell in IT to various stimuli. (**A**) (1) A naturally colored monkey face, (2) the same face with scrambled components, (3) a second monkey face, (4) the second face with snout removed, (5) eyes removed, (6) uncolored, (7) a human face, and (8) a hand. (**B**) A monkey face in different degrees of rotation in depth. (From Desimone et al., 1984).

IT cells to the *components* of response-eliciting patterns. Sato et al. (1980) found that a cell responsive to a plus sign was unresponsive (and not just half as responsive) to either its vertical or horizontal component when presented in isolation. Iwai (1985) replicated these results for IT cells, and then divided foveal prestriate cells into two groups based on how their responses relate to the components of the pattern to which they maximally responded. Unlike more anterior cells, the first group was unresponsive to rotations and scalings of the pattern, as well as being unresponsive to its components. The responses of the second group appeared to be selective for a particular component of the pattern, rather than to the pattern per se, so that a rotated or scaled version of another pattern containing that component would produce as vigorous a response. Desimone et al. (1984) found face-selective cells that were unresponsive to isolated facial components and were unresponsive to faces in which the components are scrambled, demonstrating that the response of these cells depended on the spatial relations between facial components. Yamane et al.

(1988) parametrically varied the structure of face stimuli and found that face-selective cells responded to combinations of distances among different facial features.

A common assumption is that the representation of faces is typical of object representation in general (e.g., Desimone et al., 1984). However, there are at least two reasons to suspect that the mechanisms of face recognition may differ from general object recognition. The special significance of faces as visual stimuli, and the anatomical segregation of face-selective cells both suggest that our visual systems may have developed specialized kinds of representation for faces. Thus caution is warranted in generalizing from properties of face-selective cells to characteristics of object representation in general.

Summary

The three sources of evidence just reviewed all implicate IT in the highest levels of visual object representation. Results from studies of brain-damaged humans suggest

that the ventral regions of the temporal lobe are impor-
tant for perceiving the shape of objects in their entirety,
as opposed to one small portion at a time, and that
without this ability people cannot recognize objects. Re-
search with animals has confirmed the role of IT in
higher vision with more precise experimental lesions. In
addition, this research has characterized more precisely
the kinds of visual information represented by IT. Mon-
keys with lesions in this area are unable to respond to a
particular object as being the same after it has undergone
a change in location, size, contrast, lighting, or orienta-
tion. This implies that the ability to represent the shape
of an object, independent of lower level image proper-
ties, depends upon IT. Finally, recordings from single
neurons in this area provide an even finer grained char-
acterization of the kinds of information coded in IT. As
one would expect, given the results of IT ablations in
animals, many neurons in this area respond selectively
to a particular shape, roughly independent of its retinal
location, size, contrast, and picture plane orientation. In
contrast to the responses of neurons in earlier visual
areas, some neurons in IT respond selectively to whole,
complex objects such as faces and hands. The depen-
dence of the responses of these cells on the overall
spatial structure of objects is consistent with the behavior
of human associative agnosic patients, who appear un-
able to perceive whole complex objects, and with reports
that some of these patients have disproportionate diffi-
culty recognizing faces.

In the next section we will review and evaluate several
proposals that have been put forth to explain the data
just discussed. We will then consider the relation be-
tween these data and the computational issues in object
recognition previously discussed.

THEORIES OF INFEROTEMPORAL FUNCTION

Perceptual Constancy

Several proposals have been offered for the role of IT
in object recognition. Perhaps the most widely accepted
of these is that IT provides *perceptual constancy*, that is,
the ability to see that two inputs with different retinal
positions, orientations, and sizes arise from the same
physical object (Desimone et al., 1985; Gross, 1978; Gross
& Mishkin, 1977; Holmes & Gross, 1984a,b; Iwai, 1985;
Laursen, 1982; Seacord et al., 1979). On this view, the
discrimination deficit following IT lesions is due to the
fact that successive presentations of the target stimulus
have slightly different retinal projections, and the mon-
key lacks the mechanism that normally indicates that
these stimuli represent the same object. Thus the monkey
is faced with learning a large number of separate dis-
criminations between each apparently different target
object and the distractors.

Although this interpretation is certainly consistent with
many of the results from lesion studies and single-cell
recordings, is is little more than a redescription of these
results in terms of the well-established psychological
term "constancy." It fails to extend our understanding or
generate more precise predictions. In particular, it tells
nothing of how IT subserves the class of abilities referred
to as perceptual constancy.

Categorization

Dean (1982) suggested that what is stored in visual mem-
ory is a simplified, impoverished description of the rich
perceptual input. Dean referred to the process of deriv-
ing this briefer, more symbolic description of lower level
visual information as *categorization,* and hypothesized
it as the role of IT in high-level vision. This explanation
is consistent with the claim that IT subserves perceptual
constancy because ignoring changes in viewpoint may
be part of the process of deriving the simplified descrip-
tion. IT lesions eliminate preoperatively learned discrim-
inations by destroying the description of the target that
was associated with reward. The postoperative learning
deficit arises because, without the normal mechanism for
describing the stimuli, the monkey must rely on lower
level, less succinct descriptions.

Unfortunately, the notion of "categorization" is also
insufficiently precise to generate interesting experimen-
tal predictions. In fact, the notion is so imprecise that
both of two diametrically opposed versions of the hy-
pothesis are consistent with existing results. Assuming
that the categorization of monkeys with IT lesions is
impaired rather than eliminated, their overgeneralization
to similar stimuli (Butter et al., 1965) suggests that their
categorization is abnormally *imprecise.* On the other
hand, the fact that these monkeys show reduced transfer
to transformed versions of the discrimination target
(Weiskrantz & Saunders, 1984) suggests that the descrip-
tions they are using are overly *precise,* in that they take
into account information that depends on viewpoint. The
explanatory usefulness of the notion of "categorization"
is questionable given that it must be used in such differ-
ent and conflicting ways to account for the data. Fur-
thermore, the exact nature of the "symbolic" description
and its derivation remains unspecified.

Distributed-Trace Memory

Gaffan et al. (1986a,b) proposed that IT functions as a
"distributed-trace" memory (Anderson, 1973; Hinton &
Anderson, 1981), in which stimuli are represented as a
long list of values of visual attributes. Typically, each
possible attribute value is represented by a separate neu-
ron-like processing unit, so that the representation of a

stimulus consists of a pattern of activity over these units (see the discussion of distributed representations). Associating a stimulus with reward during discrimination learning amounts to associating each active attribute unit with the reward, which can be accomplished by increasing the weights on connections between active units. The number of associations that can be stored without interference in such a system increases with the extent to which the stimuli are dissimilar, and the number of attributes available to describe stimuli. Gaffan and his colleagues explain the discrimination deficit following IT lesions as being the result of a decrease in the number of attributes input to the distributed-trace memory system.

This proposal is consistent with existing data on deficits following IT lesion, and is appealing in that it is more neurally explicit than other explanations. Also, the proposed *functional* deficit (fewer descriptive attributes) corresponds directly with the known *anatomical* deficit (cortical lesion) under the plausible assumption that neurons represent attribute values. This natural correspondence is a consequence of using a neural-like computational architecture rather than one based on conventional symbol-manipulation. However, in the framework previously presented, this interpretation of IT function is explicit about the *implementation* of IT representations, without being very explicit about the representations themselves. Detailed predictions are difficult to derive without a more precise specification of the nature of the attributes actually used to describe objects.

Object-Centered Prototypes

Ratcliff and Newcombe (1982) made a more specific proposal about the form of the descriptions underlying object recognition, thereby providing an elaboration of, rather than an alternative to, Dean's categorization hypothesis. They suggested that agnosic patients have lost the ability to construct object representations akin to Marr and Nishihara's (1978) object-centered 3D models. Weiskrantz and Saunders (1984) made a similar proposal, suggesting that IT is the locus for storing an object-centered "prototype" of a visual object in a form that is accessed by visual information from translated, rotated, or scaled versions of the object. More posterior cortical regions, including foveal prestriate cortex, are hypothesized to represent visual information in a viewer-centered format, and their anterior projections are involved in addressing the object-centered prototype based on this viewer-centered information. In visual discrimination tasks, IT lesions force the monkey to rely on viewer-centered information, which provides a description with which to associate reward that is less complete and precise than the object-centered descriptions that normal monkeys use.

The hypothesis that IT contains object-centered prototypes that are addressed by viewer-centered descriptions in foveal prestriate cortex is the most complete, predictive existing explanation of the role of IT in object recognition. It goes beyond previous explanations by attempting to specify the types of representations involved in recognition, and the nature of the processes that operate over these representations. Yet it is incomplete in that it fails to specify properties of the prototypes themselves, beyond claiming that they are object-centered. Also, as was previously pointed out, the extent to which a reference frame is object-centered can be more a matter of degree than of kind, so the claim that object representations are object-centered is underspecified.

A Computational Interpretation of Inferotemporal Function

In this section we attempt to characterize object representations more precisely by interpreting the experimental results on representations in IT in terms of the computational issues previously discussed. Since any representation can mimic any other by employing additional processes (assuming no information loss), the nature of a representation can never be uniquely determined independent of the broader processing context (Anderson, 1978). Accordingly, our conclusions are limited to the form: the available physiological data are more consistent with particular types of representations, interpreted in the most straightforward way (i.e., without postulating compensatory mechanisms). Thus, as much as we would like to be able to specify the nature of these representations definitively, the conclusions that can be drawn from the available evidence must be viewed as tentative. Yet even tentative relationships between theory and data can usefully guide further investigation and constrain existing models.

Shape Primitives

The main difference between contour-based primitives on the one hand, and surface-based or volumetric primitives on the other, is that the former are less stable than the latter. Thus deriving a contour-based object representation that is stable in the face of confounding image variation (e.g., changes in viewpoint or lighting) requires additional and/or more complicated mechanisms to compensate for the relative instability of the primitives. Hence, the available neurophysiological data would be more difficult to account for in terms of a completely contour-based representation, as compared with either a surface-based or volumetric representation. Yet in considering the data it is important to keep in mind that

none of these alternatives can be strictly ruled out on the basis of existing evidence.

Brain-damaged agnosic patients appear to have difficulty seeing stimuli in terms of surfaces and volumes. This is suggested by their poor performance on matching tasks in which objects must be matched across changes in perspective, for example the version of Warrington's unusual views task that was given to Ratcliff and Newcombe's (1982) subject M.S., and the Benton and Van Allen face matching task. Producing identical representations of an object across changes in perspective would be easier using surfaces or volumes than using two-dimentional contours. Similarly, the changes in illumination across the faces to be matched in one section of the Benton and Van Allen face matching task would also result in more drastic changes in a contour-based representation, as the shadows are more likely to be misinterpreted as relevant contours than as surfaces or volumes. The poor performance of agnosic patients on these tasks suggests that they are overly distracted by the additional contours because they can no longer generate more stable surface and/or volumetric representations. Ratcliff and Newcombe's demonstration that M.S. cannot discriminate possible from impossible figures is also relevant to the issue of primitives. Although there are computational systems that can distinguish between these types of figure on the basis of contour information such as junctions (Waltz, 1975), the definitions of what constitute illegal adjacent junction combinations implicitly assume a surface or volumetric interpretation. That M.S. could not make this type of discrimination is consistent with his inability to derive such an interpretation. Finally, to the extent that copying strategies reveal properties of the underlying visual representation, the slavish, line by line copying strategies of these patients also suggest that they are relying more on local contour in their copying than would a normal person.

IT-lesioned monkeys show a similar reliance on local contour information, and an inability to see the equivalence of three-dimensional stimuli that have undergone changes in lighting or perspective (Weiskrantz & Saunders, 1984). Again, this suggests that they are relying on representations that are neither surface-based nor volumetric, and that the normal function of IT must therefore include the representation of shape using either surface or volumetric primitives (or both). More direct evidence that surface representations are computed in IT comes from the experiment of Cowey and Porter (1979), which showed that IT-lesioned monkeys were impaired at perceiving surfaces in depth in random dot stereograms, which do not have contours.

Recordings of single IT neurons generally reveal greater responses to three-dimensional objects than to drawings (Desimone et al., 1984). Assuming that line drawings capture the essential contours of the object they are depicting, this result can be taken as further con-

verging evidence that IT cortex normally represents shape in terms of either surface or volumetric primitives.

Reference Frame

Previous discussions of frame of reference in visual neurophysiology have distinguished between the general concepts of viewer-centered and object-centered frames (Perrett et al., 1985; Ratcliff & Newcombe, 1982; Weiskrantz & Saunders, 1984). However, depending on the data being considered, different conclusions seem to be implied. The invariance of single unit responses to objects over transformations of location, size, and image plane orientation suggests that temporal cortex houses object-centered representations of shape. In contrast, the sensitivity of face cells to depth orientation implies a viewer-centered frame. One way to reconcile these different findings and interpretations is to suppose that both types of representation are used in temporal cortex (cf. Weiskrantz & Saunders, 1984). However, it is also possible that temporal object representations are object-centered wth respect to certain of their degrees of freedom, and viewer-centered with respect to others (see previous for a discussion of degrees of freedom in reference frames). This latter interpretation is more consistent with the finding that a given cell may have orientation-invariant responses for image-plane rotations, but not for rotations in depth. The available data suggest that while position, scale, and image-plane orientation are object-centered, orientation in depth seems at least partially viewer-centered. This pattern of results is generally consistent with an object representation based on "characteristic-views" (Ikeuchi, 1987; Koenderink & van Doorn, 1979). In addition, the reference frame does not appear to make the handedness of an object explicit.

In support of these conclusions, human agnosics have difficulty seeing the equivalence of objects across changes in depth orientation. In the Benton and Van Allen face-matching task, agnosics have more difficulty than normals matching across changes in perspective than when matching identical views. Also, Ratcliff and Newcombe's (1982) agnosic patient M.S. is unable to relate usual and unusual views of objects. As a result of having lost their object-centered representations, these patients have difficulty seeing the equivalence of objects across depth rotations.

IT-lesioned monkeys also do poorly at seeing the equivalence of three-dimensional shapes when viewed from different perspectives, implying that IT normally represents objects in such a way that different views of an object map onto the same (object-centered) representation. These monkeys have difficultly generalizing a learned discrimination to versions of the target object rotated in depth (Weiskrantz & Saunders, 1984). They also show less "constancy interference" than normals in

discrimination tasks: IT-lesioned monkeys are less bothered when the discriminanda differ only in orientation (when constancy interferes with discrimination). In addition, whereas normal monkeys find image-plane rotated-pattern discriminations more difficult than different-pattern discriminations, this difference is much smaller for monkeys with IT lesions, implying that IT represents the rotated versions of a pattern as equivalent. In a similar way, patterns that differ only in handedness are more difficult for normals to discriminate (Gross et al., 1975), suggesting that enantiomorphs are given equivalent descriptions in IT.

The response properties of cells in IT are relatively unaffected by changes in the retinal position, size, and image-plane orientation of stimuli. However, face-selective cells in IT (and STS) have been found that are selective for particular orientations in depth (Desimone et al., 1984; Perrett et al., 1985). Note that these results appear to conflict with those of Weiskrantz and Saunders (1984) described above, showing that IT is important for generalizing across depth rotations of ordinary objects. However, it is important to keep in mind that the particular frequency, featural configuration, and relevance of facial stimuli for monkeys (and humans) may have resulted in the development of more special-purpose representations for these stimuli whose properties may not apply to the representation used for general objects. In summary, these results suggest the use of a reference frame that is object-centered along all dimensions except (at least for faces) orientations in depth.

Organization

Compared with the previous two issues, the neurophysiological data have little to say about the nature of the organization of object representations in IT, although they are consistent with much computational and psychological work (Biederman, 1987; Hoffman & Richards, 1985; Marr & Nishihara, 1978; Palmer, 1977; Pomerantz, Sager, & Stoever, 1977) that suggests that objects are decomposed into parts and their spatial relationships.

Human agnosics do poorly on tasks that require the explicit representation of parts and their relations to one another. They have difficult distinguishing possible from impossible objects (Ratcliff & Newcombe, 1982), a task that involves verifying the global consistency of locally consistent parts. Agnosic L.H. was better than normals at seeing "bad" parts in embedded figures (Levine & Calvanio, 1989), presumably because he was less susceptible to interference from decompositions into "good" parts.

Iwai (1985) found IT cells that were selective for particular components of a simple pattern. Desimone et al. (1984) found face-selective cells whose responses were eliminated by spatially rearranging the components of a face, and that were unresponsive to individual compo-

nents. Again, caution is warranted in generalizing the results on the representation of faces to the representation of objects in general. Although determining the *class* of an object is sufficient for most object recognition tasks, face recognition usually involved identifying an *individual,* which may require a more precise metrical representation of the spatial relationships of parts (Bruce, 1988). Therefore the available data do not indicate the extent to which the spatial organization of parts of objects other than faces is explicitly represented.

Implementational Issues

The finding that certain cells respond selectively to particular stimuli might at first seem to imply local representations of the "grandmother cell" variety, as opposed to distributed representations. However, given the frequency with which a randomly selected cell responds to one of the stimuli selected by the experimenter in single-unit recording studies, it seems clear that a large population of cells is to some degree responsive during the recognition of any stimulus. Furthermore, even highly selective cells, such as those that respond differentially to different faces, will respond to a range of stimuli. On the basis of these observations, it seems likely that temporal neurons represent objects in a distributed manner, with different portions of the population being active to different degrees depending on the stimulus. (Direct evidence for such a system of representation has been found in the motor system by Georgopoulos, Schwartz, and Kettner, (1986) using multiple simultaneous single unit recordings.) Another aspect of the single-unit data that is consistent with the notion of distributed, rather than local, representation concerns the degree of constancy over transformations in stimulus position, size, and image-plane orientation of single cells. Although single cells do show shape selectivity over a wide range of locations, for example, they respond most strongly within a certain subset of those locations (e.g., see Desimone et al., 1984). Thus the responses of individual cells are not as invariant as the behavior of the animal. The degree of shape constancy displayed behaviorally is presumably the result of a population of such neurons, with overlapping receptive fields, responding together.

Another clue to the implementation of object recognition comes from the study of human agnosia. Traditionally, associative agnosia was interpreted as a loss of stored visual memories, with intact perception. This conception of agnosia is consistent with a symbolic architecture, in which a representation derived from the stimulus during perception is matched against a separate stored representation. In contrast, in a connectionist architecture, the ability to derive the final perceptual representation depends on the "memories" that are encoded in the connection strengths. In such an architecture, it

would be impossible to have damaged memory with intact perception. As noted earlier, in cases of associative agnosia in which perception has been tested carefully, it has been found to be impaired. The lack of dissociability between perception and memory for objects in agnosia is therefore consistent with a connectionist implementation of object recognition.

CONCLUSION

A large body of neurophysiological data shows that inferior temporal cortex plays a critical role in the representation and recognition of visual objects. Cells in IT respond selectively to physical properties of distal objects rather than the more variable properties of the proximal image, and damage to this area in humans as well as in monkeys produces systematic deficits in visual recognition and discrimination of objects and disproportionate reliance on proximal cues in visual tasks. Despite the wealth of data implicating IT in object representation, theories about the function of IT have been slow to emerge and have not played a dominant role in directing on-going research. What is needed are theories of IT function that are sufficiently precise to account for existing data and to generate specific, testable predictions.

Computational research on object recognition is concerned with analyzing the problems faced by *any* visual recognition system, and in determining how these problems can be solved given the available information. Furthermore, computational vision researchers have developed a set of explicit distinctions necessary for precise theorizing about object representation. Thus the ideas of computational vision are both relevant to the neurophysiology of object recognition and potentially useful for casting theories of IT function in more precise, testable ways. In this paper we have described a computational framework in which specific questions can be posed about the nature of object representation, and we have interpreted the existing neurophysiological data on object representation in terms of their implications for answering these questions.

We have not proposed an alternative *theory* of IT function. Rather, we have pointed out some important theoretical distinctions about the computational problem of object recognition that should promote the development of more precise theories of IT function. Casting existing neurophysiological data in terms of computational distinctions serves both to suggest particularly informative experimental issues (e.g., whether shape primitives are surface or volume-based, and the extent to which object representations involve fully or only partially object-centered representations) and to generate more applicable empirical constraints on computational models of object recognition (e.g., that contour-based primitives appear insufficient). This kind of interdisciplinary interaction has proven valuable in the study of low-level vision—we believe it also can be of value in the study of high-level vision.

References

Alexander, M. P., & Albert, M. L. (1983). The anatomical basis of visual agnosia. In A. Kertesz, (Ed.), *Localization in neuropsychology*. Academic Press. New York: pp. 393–415.

Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review, 80,* 417–438.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85,* 249–277.

Badler, N., & Bajcsy, R. (1978). Three-dimensional representations for computer graphics and computer vision. *Computer Graphics, 12,* 153–160.

Bajcsy, R., & Solina, F. (1987). Three-dimensional object representation revisited. In *Proceedings, 1st International Conference on Computer Vision*, pp. 231–240, London, England.

Ballard, D. H., & Brown, C. M. (1982). *Computer vision*. Englewood Cliffs, New Jersey: Prentice Hall.

Ballard, D. H., Hinton, G. E., & Sejnowski, T. J. (1983). Parallel visual computation. *Nature (London), 306,* 21–26.

Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex of the superior temporal sulcus of the monkey. *Brain Research, 342,* 91–102.

Benson, D. F., & Greenberg, J. P. (1969). Visual form agnosia. *Archives of Neurology, 20,* 82–89.

Benton, A. L., & Van Allen, M. W. (1972). Prosopagnosia and facial discrimination. *Journal of Neurological Sciences, 15,* 167–172.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Blake, L., Jarvis, C. D., and Mishkin, M. (1977). Pattern discrimination thresholds after partial inferior temporal or lateral striate lesions in monkeys. *Brain Research, 120,* 209–220.

Blum, H. (1973). Biological shape and visual science (Part I). *Journal of Theoretical Biology, 38,* 205–287.

Blum, J. S., Chow, K. L., & Pribram, K. H. (1950). A behavioral analysis of the organization of the parieto-temporo-preoccipital cortex. *Journal of Comparative Neurology, 93,* 53–100.

Brooks, R. A. (1981). Symbolic reasoning among 3-D models and 2-D images. *Artificial Intelligence, 17,* 285–348.

Brown, T. S. (1963). Olfactory and visual discrimination in the monkey after selective lesions of the temporal lobe. *Journal of Comparative and Physiological Psychology, 56,* 764–768.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology, 46(2),* 369–384.

Bruce, V. (1988). *Recognizing faces.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Butter, C. M. (1968). The effect of discrimination training on pattern equivalence in monkeys with infero-temporal and lateral striate lesions. *Neuropsychologia, 6,* 27–40.

Butter, C. M., Mishkin, M., & Rosvold, H. E. (1965). Stimulus generalization in monkeys with infero-temporal lesions and lateral occipital lesions. In D. J. Mostofsky, (Ed.), *Stimulus generalization.* Stanford: Stanford University Press. pp. 119–133.

Chow, K. L. (1951). Effects of partial extirpations of the posterior association cortex on visually mediated behavior. *Comparative Psychology Monographs, 20,* 187–217.

Chow, K. L. (1952). Further studies on selective ablation of associative cortex in relation to visually mediated behavior. *Journal of Comparative and Physiological Psychology, 45,* 109–118.

Churland, P. S., & Sejnowski, T. J. (1988). Perspectives on cognitive neuroscience. *Science, 242,* 741–745.

Cowey, A. & Gross, C. G. (1970). Effects of foveal prestriate and inferotemporal lesions on visual discrimination by rhesus monkeys. *Experimental Brain Research, 11,* 128–144.

Cowey, A. and Porter, J. (1979). Brain damage and global stereopsis. *Proceedings, Royal Society of London, Series B, 204,* 399–407.

Cowey, A. & Weiskrantz, L. (1967). A comparison of the effects of inferotemporal and striate cortex lesions on the visual behavior of rhesus monkeys. *Quarterly Journal of Experimental Psychology, 15,* 91–115.

Dean, P. (1974). The effect of inferotemporal lesions on memory for visual stimuli in rhesus monkeys. *Brain Research, 77,* 451–469.

Dean, P. (1982). Visual behavior in monkeys with inferotemporal lesions. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield, (Ed.), *Analysis of visual behavior.* Cambridge, Massachusetts: MIT Press. pp. 587–628.

Dean, P., & Cowey, A. (1977). Inferotemporal lesions and memory for pattern discriminations after visual interference. *Neuropsychologia, 15,* 93–98.

Dean, P., & Weiskrantz, L. (1977). Loss of preoperative habits in rhesus monkeys with inferotemporal lesions: Recognition failure or relearning deficit? *Neuropsychologia, 12,* 299–311.

Derthick, M., & Plaut, D. C. (1986). Is distributed connectionism compatible with the Physical Symbol System Hypothesis? In *Proceedings, 8th Annual Conference of the Cognitive Science Society,* pp. 639–644, Amherst, Massachusetts.

Desimone, R., & Gross, C. G. (1979). Visual areas in the temporal cortex of the macaque. *Brain Research, 178,* 363–380.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience, 4,* 2051–2062.

Desimone, R., Schein, S. J., Moran, J., & Underleider, L. G. (1985). Contour, color and shape analysis beyond the striate cortex. *Vision Research, 25(3),* 441–452.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, New Jersey: Educational Testing Service.

Ellis, A. W. & Young, A. W. (1987). *Human cognitive neuropsychology.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision.* Cambridge, Massachusetts: MIT Press.

Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology, 20,* 439–462.

Feldman, J. A. & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science, 6,* 205–254.

Fuster, J. M. & Jervey, J. P. (1982). Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *Journal of Neuroscience, 2,* 361–375.

Gaffan, D., Harrison, S., & Gaffan, E. A. (1986a). Visual identification following inferotemporal ablation in the monkey. *Quarterly Journal of Experimental Psychology, 38B,* 5–30.

Gaffan, D., Harrison, S., & Gaffan, E. A. (1986b). Single and concurrent discrimination learning by monkeys after lesions of inferotemporal cortex. *Quarterly Journal of Experimental Psychology, 38B,* 31–51.

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science, 233,* 1416–1419.

Gregory, R. L. (1970). *The intelligent eye.* New York: McGraw-Hill.

Gross, C. G. (1973). Inferotemporal cortex and vision. In E. Stellar, & J. M. Sprague, (Eds.), *Progress in Physiological Psychology,* New York: Academic Press.

Gross, C. G. (1978). Inferior temporal lesions do not impair discrimination of rotated patterns in monkeys. *Journal of Comparative and Physiological Psychology, 92,* 1095–1109.

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science, 166,* 1303–1306.

Gross, C. G., Cowey, A., & Manning, F. J. (1971). Further analysis of the visual discrimination deficits following foveal prestriate and infero-temporal lesions in rhesus monkeys. *Journal of Comparative and Physiological Psychology, 76(1),* 1–7.

Gross, C. G., Lewis, M., & Plaisier, D. (1975). Inferior temporal cortex lesions do not impair discrimination of lateral mirror images. *Society for Neuroscience Abstracts, 1.*

Gross, C. G. & Mishkin, M. (1977). The neural basis of stimulus equivalence across retinal translation. In S. Harnard, R. W. Doty, L. Goldstein, J. Jaynes, & G. Krauthamer, (Eds.), *Lateralization in the nervous system,* New York: Academic Press.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology, 35,* 96–111.

Gross, C. G., Schiller, P. H., Wells, C., & Gerstein, G. L. (1967). Single unit activity in temporal association cortex of the monkey. *Journal of Neurophysiology, 30,* 833–843.

Heywood, C. A., & Cowey, A. (1987). On the role of cortical area V4 in the discrimination of hue and pattern in macaque monkeys. *Journal of Neuroscience, 7(9),* 2601–2617.

Hinton, G. E., & Anderson, J. A. (1981). *Parallel models of associative memory.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Hinton, G. E., & Parsons, L. M. (1988). Scene-based and viewer-centered representations for comparing shapes. *Cognition, 30,* 1–35.

Hinton, G. E., & Sejnowski, T. J. (1983). Analyzing cooperative

computation. In *Proceedings, 5th Annual Conference of the Cognitive Science Society*, Rochester, New York.

Hinton, G. E. & Sejnowski, T. J. (1986). Learning and re-learning in Boltzmann machines. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, Massachusetts: MIT Press. pp. 282–317.

Hinton, G. E. & Shallice, T. (1990). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review* (in press).

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group. (Eds) *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, Massachusetts: MIT Press. pp. 77–109.

Hoffman, D. D., & Richards, W. A. (1985). Parts of recognition. *Cognition, 18,* 65–96.

Holmes, E. J., & Gross, C. G. (1984a). Effects of inferior temporal lesions on discrimination of stimuli differing in orientation. *Journal of Neuroscience, 4(12),* 3063–3068.

Holmes, E. J., & Gross, C. G. (1984b). Stimulus equivalence after inferior temporal lesions in monkeys. *Behavioral Neuroscience, 98(5),* 898–901.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings, National Academy of Science, U.S.A., 79,* 2554–2558.

Horn, B. K. P. (1986). *Robot vision.* Cambridge, Massachusetts: MIT Press.

Humphrey, N. K., & Weiskrantz, L. (1969). Size constancy in monkeys with inferotemporal lesions. *Quarterly Journal of Experimental Psychology, 21,* 225–238.

Humphreys, G. W., & Riddoch, M. J. (1987). *To see but not to see: A case-study of visual agnosia.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Huttenlocher, D. P. (1988). Three-dimensional recognition of solid objects from a two-dimensional image. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, M.I.T.

Ikeuchi, K. (1987). Generating an interpretation tree from a CAD model for 3D-object recognition in bin-picking tasks. *International Journal of Computer Vision, 1,* 145–165.

Iversen, S. D. (1970). Interference and inferotemporal memory deficits. *Brain Research, 19,* 227–289.

Iwai, E. (1985). Neuropsychological basis of pattern vision in macaque monkeys. *Vision Research, 25(3),* 425–439.

Iwai, E. & Mishkin, M. (1968). Two visual foci in the temporal lobe of monkeys. In N. Yoshii, & N.A. Buchwald, (Eds.), *Neurophysiological basis of learning and behavior.* Osaka: Osaka University Press.

Iwai, E. & Mishkin, M. (1969). Further evidence of the locus of the visual area in the temporal lobe of the monkey. *Experimental Neurology, 25,* 585–594.

Kanade, T. (1987). Three-dimensional machine vision. Boston: Kluwer Academic Publishers.

Kluver, H. & Bucy, P. C. (1937). "Psychic blindness" and other symptoms following bilateral temporal lobectomy in rhesus monkeys. *American Journal of Physiology, 119,* 352–353.

Kluver, H., & Bucy, P. C. (1939). Preliminary analysis of functions of the temporal lobes of monkeys. *Archives of Neurology and Psychiatry, 42,* 979–1000.

Koenderink, J. J. & van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics, 32,* 211–216.

Kuypers, H. G. J. M., Szwarcbart, M. K., Mishkin, M., & Ros-

vold, H. E. (1965). Occipitotemporal corticocortical connections in the rhesus monkey. *Experimental Neurology, 11,* 245–262.

Laursen, A. M. (1982). A lasting impairment in circle-ellipse discrimination after inferotemporal lesions in monkeys. *Behavioral Brain Research, 6,* 201–212.

Lennie, P. (1980). Parallel visual pathways: A review. *Vision Research, 20,* 561–594.

Levine, D. N. (1982). *Visual agnosia in monkey and in man* (Chapter 20). Cambridge, Massachusetts: MIT Press. pp. 629–670.

Levine, D. N. & Calvanio, R. (1989). Prosopagnosia: A defect in visual-configurational processing. *Brain and Cognition, 10,* 149–170.

Lindsay, P. H. & Norman, D. A. (1977). *Human information processing: An introduction to psychology.* New York: Academic Press.

Livingstone, M. S., & Hubel, D. H. (1984). Specificity of intrinsic connections in primate primary visual cortex. *Journal of Neuroscience, 4,* 2830–2835.

Lowe, D. G. (1987). *Perceptual organization and visual recognition.* Boston: Kluwer.

Lund, J. S. (1988). Anatomical organization of macaque monkey striate visual cortex. *Annual Review of Neuroscience, 11,* 253–288.

Marr, D. (1977). Analysis of occluding contour. *Proceedings, Royal Society of London, Series B, 197,* 441–475.

Marr, D. (1982). *Vision.* San Francisco: W. H. Freeman.

Marr, D., & Hildreth, E. K. (1980). Theory of edge detection. *Proceedings, Royal Society of London, Series B, 207,* 187–217.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings, Royal Society of London, Series B, 200,* 269–294.

Maunsell, J. H. R. & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience, 10,* 363–401.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.* Cambridge, Massachusetts: MIT Press.

Meyer, D. R., Treichler, F. R., & Meyer, P. M. (1965). Discrete-trial training techniques and stimulus variables. In A. M. Schrier, H. F. Harlow, & F. Stollnitz, (Eds.), *Behavior of nonhuman primates.* New York: Academic Press. pp. 1–49.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Mishkin, M. (1954). Visual discrimination performance following partial abalations of the temporal lobe: II. Ventral surface vs. hippocampus. *Journal of Comparative and Physiological Psychology, 47,* 187–193.

Mishkin, M. (1966). Visual mechanisms beyond the striate cortex. In R. Russel, (Ed.), *Frontiers in physiological psychology.* New York: Academic Press. pp. 93–119.

Mishkin, M. (1972). Cortical visual areas and their interaction. In A. G. Karezmar, & J.C. Eccles, (Eds.), *Brain and human behavior.* New York: Springer-Verlag.

Mishkin, M. & Pribram, K. H. (1954). Visual discrimination performance following partial ablations of the temporal lobe: i. ventral vs. lateral. *Journal of Comparative and Physiological Psychology, 47,* 14–20.

Mishkin, M. & Weiskrantz, L. (1959). Effects of cortical lesions in monkeys on critical fusion frequency. *Journal of Comparative and Physiological Psychology, 52,* 660–666.

Miyashita, Y., & Chang, H. (1988). Neuronal correlate of picto-

rial short-term memory in the primate temporal cortex. *Nature (London), 331,* 68–70.

Nevatia, R., & Binford, T. O. (1977). Description and recognition of curved objects. *Artificial Intelligence, 8,* 77–98.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4,* 135–183.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9,* 441–474.

Patterson, K. E., Seidenberg, M. S., & McClelland, J. L. (1990). Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In R. G. M. Morris (Ed.), Parallel distributed processing: implications for psychology and neuroscience. London: Oxford University Press.

Pentland, A. P. (1986). Perceptual organization and the representation of natural form. *Artificial Intelligence, 28,* 293–331.

Perrett, D. I., Rolls, E. T., & Caan, W. (1979). Temporal lobe cells of the monkey with visual responses selective for faces. *Neuroscience Letters, S3,* S358.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research, 47,* 329–342.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings, Royal Society of London, Series B, 223,* 293–317.

Pinker, S. (1985). *Visual cognition.* Cambridge, Massachusetts: MIT Press.

Poggio, T. (1983). Visual Algorithms. In O. J. Braddick, & A. C. Sleigh, (Eds.), *Physical and biological processing of images.* New York: Springer-Verlag. pp. 128–153.

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3(3),* 422–435.

Pribram, K. H. (1954). Toward a science of neuropsychology: Method and data. In R. A. Patton, (Ed.), *Current trends in psychology and the behavioral sciences.* Pittsburgh: University of Pittsburgh Press. pp. 115–152.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science.* Cambridge, Massachusetts: MIT Press.

Ratcliff, G., & Newcombe, F. A. (1982). *Object recognition: Some deductions from the clinical evidence.* New York: Academic Press. pp. 147–171.

Reed, S. K. (1982). *Cognition: Theory and applications.* Monterey, California: Brooks/Cole.

Richards, W., & Hoffman, D. D. (1985). Codon constraints on closed 2D shapes. *Computer Vision, Graphics, and Image Processing, 31(2),* 156–177.

Richmond, B. J., & Sato, T. (1987). Enhancement of inferior temporal neurons during visual discrimination. *Journal of Neurophysiology, 58(6),* 1292.

Richmond, B. J., Optican, L. M., Podell, M., & Spitzer, H. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *Journal of Neurophysiology, 57(1),* 132–146.

Richter, J., & Ullman, S. (1986). *Biological Cybernetics, 53,* 195–202.

Riddoch, M. J. & Humphreys, G. W. (1987). Visual object processing in optic aphasia: A case of semantic access agnosia. *Cognitive Neuropsychology, 4(2),* 131–185.

Roberts, L. G. (1965). Machine perception of three-dimensional solids. In J. L. Tippet, (Ed.), *Optical and electro opti-*

cal information processing. Cambridge, Massachusetts: MIT Press.

Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Human Neurobiology, 3,* 209–222.

Rolls, E. T., & Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Experimental Brain Research, 65,* 38–48.

Rolls, E. T., Judge, S. J., & Sanghera, M. K. (1977). Activity of neurons in the inferotemporal cortex of the alert monkey. *Brain Research, 130,* 229–238.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, Massachusetts: MIT Press.

Sato, T., Kawamura, T., & Iwai, E. (1980). Responsiveness of inferotemporal single units to visual pattern stimuli in monkeys performing discriminations. *Experimental Brain Research, 38,* 313–319.

Schwartz, E. L., Desimone, R., Albright, T. D., & Gross, C. G. (1983). Shape recognition and inferior temporal neurons. *Proceedings, National Academy of Science, U.S.A., 80,* 5776–5778.

Seacord, L., Gross, C. G., & Mishkin, M. (1979). Role of inferior temporal cortex in interhemispheric transfer. *Brain Research, 167,* 259–272.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, England: Cambridge University Press.

Shallice, T., & Jackson, M. (1988). Lissauer on agnosia. *Cognitive Neuropsychology, 5(2),* 153–192.

Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science, 240,* 338–340.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognition, 21,* 233–282.

Ungerleider, L. G., Ganz, L., & Pribram, K. H. (1969). Size constancy in rhesus monkeys: Effects of pulvinar, prestriate, and inferotemporal lesions. *Experimental Brain Research, 27,* 251–269.

Van Essen, D. C. (1985). Functional organization of primate visual cortex. In A. Peters, & E. B. Jones, (Eds.), *Cerebral cortex.* New York: Plenum Press. pp. 259–329.

von Bonin, G., & Bailey, P. (1947). *The neocortex of Macaca mulatta.* Urbana: Univeristy of Illinois Press.

Waltz, D. A. (1975). Generating semantic descriptions from drawings of scenes with shadows. In P. H. Winston, (Ed.), *The psychology of computer vision.* New York: McGraw-Hill.

Warrington, E. K. (1982). Neuropsychological studies of object recognition. *Proceedings, Royal Society of London, Series B, 298,* 15–33.

Weiskrantz, L. & Cowey, A. (1963). Striate cortex lesions and visual acuity of the rhesus monkey. *Journal of Comparative and Physiological Psychology, 56,* 225–231.

Weiskrantz, L. & Mishkin, M. (1958). Effects of temporal and frontal cortical lesions on auditory discrimination in monkeys. *Brain, 81,* 406–414.

Weiskrantz, L. & Saunders, R. C. (1984). Impairments of visual object transforms in monkeys. *Brain, 107,* 1033–1072.

Wilson, M. (1957). Effects of circumscribed cortical lesions upon somesthetic and visual discrimination in the monkey. *Journal of Comparative and Physiological Psychology, 50,* 630–635.

Wilson, M., & Kaufman, H. M. (1969). Effect of inferotemporal lesions upon processing of visual information in monkeys. *Journal of Comparative and Physiological Psychology, 69,* 44–48.

Wilson, M., Kaufman, H. M., Zieler, R. E., & Leib, J. P. (1972). Visual identification and memory in monkeys with circumscribed inferotemporal lesions. *Journal of Comparative and Physiological Psychology, 78,* 173–183.

Wilson, W. A. & Mishkin, M. (1959). Comparison of the effects of inferotemporal and lateral occipital lesions on visually guided behavior in monkeys. *Journal of Comparative and Physiological Psychology, 52,* 10–18.

Wood, C. C. (1978). Variations on a theme by Lashley: Lesion experiments on the neural model of Anderson, Silverstein, Ritz, and Jones. *Psychological Review, 85,* 582–591.

Yamane, S., & Kaji, S., & Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? *Experimental Brain Research, 73,* 209–214.

Zeki, S. M. (1978). Functional specialization in the visual cortex of the rhesus monkey. *Nature (London), 274,* 423.