Psychobiology
1998, 26 (4), 281-308

Visual object representation: An introduction

SHAUN P. VECERA
University of Iowa, Iowa City, Iowa

What are the computational, behavioral, and neural mechanisms that give rise to object perception? In this review, I present a cognitive neuroscience overview of the literature on object representation. Marr's (1982) framework for studying complex tasks is used as a guide for the review. This framework involves analyzing a problem on three levels: (1) the computational theory, which asks what is computed and how; (2) the representation and algorithm, which focus on the representations and processes that underlie a computation; and (3) the hardware implementation, which deals with the implementation of the representations and processes. Computational considerations of object recognition raise the importance of the *object invariances*, which allow viewers to perceive an object as remaining stable despite changes in the retinal image. I then use the invariances to guide my review of the representations and processes involved in human object recognition, Marr's second level, and of the hardware implementation, Marr's third level. Throughout the review, my focus is on integrating across disciplines and across the levels of Marr's framework.

Anyone who has considered how the visual system takes the retinal image and, somehow, represents and recognizes objects appearing in this retinal array has contemplated the complexity of object representation. But it is trivial to say that object representation is a complex visual task and to leave the analysis at that. The more important, and more interesting, task is to try to arrive at an understanding of the mechanisms that the visual system uses in order to represent objects. Although this task has been taken up by researchers in many different fields, this research has been occasionally disjointed and unsystematic because of the failure to relate discoveries across different disciplines. It has been only recently, with the emergence of cognitive neuroscience, that the problem of object representation has been addressed in a more systematic way, using a variety of converging methodologies that link results across both disciplines and methodologies.

The foundations of an interdisciplinary approach to object representation can be seen clearly in Marr's (1982) seminal book. In his first chapter, Marr discusses how understanding a complex task, such as object recognition, requires an analysis at multiple levels. Marr's first level is that of the *computational theory*; this requires asking about the goal of a particular computation and how this computation could be carried out. The second level of analysis, *representation and algorithm*, deals with the representations and processes that allow the computational theory to be implemented and carried out. The third level is that of *hardware implementation*, which addresses the physical implementation of the representations and processes. Each level can be loosely associated with a different discipline (e.g., computational theory and computer science, representation and cognitive psychology, hardware implementation and the neurosciences), but, on Marr's argument, an understanding of a problem at all levels, which would require cross-discipline interactions, would provide the most complete understanding of that problem.

This introduction serves to orient readers to relevant issues in object representation that will serve as an overview of current cross-disciplinary research on the topic. This is by no means an exhaustive review; research on object representation and recognition is so broad that a comprehensive review would be unwieldy to write and to read. I have chosen to organize my review around how visual systems are able to represent and recognize objects despite the tremendous retinal variability across images of the same object—that is, how visual systems arrive at object representations that remain invariant across changes in the retinal image. Because of my focus on object invariances, some studies on object representation are not relevant to the review and have been omitted (e.g., studies of visual mental imagery). The interested reader is referred to other reviews (e.g., see Edelman, 1997; Pinker, 1984; Plaut & Farah, 1990) to supplement the present review.

In this introduction, I first will cover computational issues pertaining to object processing, as well as specific models of object representation. This section addresses Marr's (1982) computational theory level. The second section covers behavioral results from cognitive psychology and psychophysics, a discussion that corresponds to Marr's representation and algorithm level. The final two sections address Marr's hardware implementation by discussing both neurophysiological studies from nonhuman primates and human neuropsychological and neuroimaging studies. Throughout, my emphasis is on an integration of the different levels of analysis and whether different

This paper was prepared while the author was at the University of Utah. Thanks to both Ray Kesner and Maureen Marron for discussion and comments on this review. Thanks also to Nancy Kanwisher, who contributed to the section on neuroimaging through recent discussions. Correspondence concerning this paper should be addressed to S. P. Vecera, Department of Psychology, 11 Seashore Hall E., University of Iowa, Iowa City, IA 52242-1407 (e-mail: shaun-vecera@uiowa.edu).

methodologies converge to common solutions of problems in object representation.

COMPUTATIONAL ISSUES AND MODELS

Light reflecting off of objects in the environment casts retinal images that have tremendous variability: Objects can occupy a multitude of retinal locations, cast retinal images of different sizes on the basis of their distance from the viewer, appear under different lighting conditions, and cast different retinal images on the basis of their orientation in the environment. Traditionally, computational approaches to vision have had the goal of reconstructing, representing, and describing the physical regularities present in the external world despite the variability in the retinal image. Computational accounts have also focused on how object representations can be stored in visual memory, allowing for later recognition of familiar or previously seen objects.

Object representations must ignore the variability inherent in the retinal image to allow perceivers to generalize across retinal variability. For example, if object representations were unable to ignore retinal variability, every time an object appeared in a different location that object would be perceived differently. Thus, object representations must remain invariant in the face of low-level, retinal changes. Specifically, optimal object representations should possess, at least, (1) *translation invariance* or *spatial invariance*, in which object representations are insensitive to the retinal position or spatial location that an object occupies; (2) *size invariance*, in which object representations are insensitive to the size of the retinal image, determined by the distance between the external object and the viewer; and (3) *orientation independence*, in which representations are insensitive to the orientation of the object in the external world.

The perceptual invariances are typically motivated by appealing to computational efficiency: Only one object representation needs to be stored in visual memory if that representation remains invariant with respect to position, size, and orientation changes. If an object representation does not have these invariances, multiple object representations would need to be stored in visual memory, resulting in a more computationally complex system. Although computational efficiency has been a major motivating factor for focusing on the perceptual invariances, translation, size, and orientation invariance have been studied at both Marr's (1982) representational level and the hardware level. Because the perceptual invariances are critical for efficient, robust object recognition, these factors will form a common thread that pervades this review and will cut across Marr's levels of analysis.

Computational approaches to object representation must specify algorithms that allow object representations to possess translation, size, and orientation invariance, and several different classes of models have been developed to explain object representation. I next turn to a discussion of the major types of models that have been developed (for other reviews, see Edelman, 1997, and Pinker, 1984): template models, feature models, and volumetric models.

Template Models

Template models involve a direct match from the retinal image to an object representation. Each object representation stored in visual memory is an exact memory (or template) of the pattern of retinal activation. Thus, recognition amounts to comparing a given retinal pattern with all of the templates stored in memory and then selecting the best-fitting template. This best-fitting template would reflect the object that was present in the retinal array and would allow recognition of this object. Recognition systems that rely on template representations exist and are able to perform some tasks, such as the pattern recognition performed by supermarket checkout scanners, quite well (see Anderson, 1995, and Neisser, 1967, for examples).

Despite the demonstrated usefulness of template-based vision systems in some domains, the problems with such systems are well known (see Neisser, 1967, for an early discussion) and limit their theoretical and practical usefulness. For example, template models are extraordinarily sensitive to changes in the retinal array; this violates the invariances that object representations are thought to exhibit. Moving an object slightly will alter its retinal image, thereby preventing a match with the appropriate template; the same holds for objects that appear at different sizes or orientations. Furthermore, two different objects that are similar, the classic example being a *P* and an *R*, could possibly be indistinguishable in a template system; the *R* input, for example, could activate the template for the *P*, creating a situation in which the visual system wouldn't know whether an *R* or a *P* was present.

In addition to these computational problems, template models also have problems accounting for neurophysiological results pertaining to object representation. As discussed in detail later, the receptive fields of neurons in the inferotemporal (IT) cortex tend to be very large and may provide a mechanism for translation (spatial) invariance in object representation (see, e.g., Gross & Mishkin, 1977) in which an object is represented irrespective of what location it occupies. A rigid template representation would be unable to code an object irrespective of retinal location (or code across other retinal variability). Although template models are conceptually simple and computationally easy to construct (as is evidenced by optical character recognition systems), they are too limited to explain general object representation.

Feature Models

Instead of matching directly from the retinal array to an object representation, as in template models, feature models involve the construction of an object representation by coding an object's geometry via the outputs of feature detectors. Object representations are built from these lower level feature detectors and contain informa-

tion about the image features present and the relative positioning of those features. In these models, feature detectors represent image edges (e.g., horizontal or vertical edges) or conjunctions of edges or features (e.g., T junctions and other vertices), and these feature detectors are replicated at each retinal location. For example, the letter *H* could be represented as two vertical line segments and a horizontal line segment, with the horizontal segment bisecting the two vertical segments forming two T junctions in the image.

Feature models have a long history in cognitive psychology, beginning with Selfridge and Neisser's (1960) Pandemonium model. Pandemonium involved a parallel analysis of features present in the visual field; there were, for example, detectors for vertical lines, crossbars, and so forth. On the basis of the input pattern, the feature detectors assigned a probability to the likelihood of certain features. These feature-based probabilities then could be used to assign a probability as to which object (e.g., a letter) was likely to be present in the input pattern. For example, the presence of two vertical lines and one horizontal line, with the appropriate interrelations, could allow the *H* object representation to be assigned a high probability, because there is a large amount of featural evidence for the letter *H*. The characterization of this process is one of a *feature demon* shouting whenever its corresponding feature is detected, and the loudness with which the demon shouts corresponds to the amount of evidence (the probability) for that feature in the input (see Selfridge & Neisser, 1960; see also Lindsay & Norman, 1977).

Early feature-based models such as Pandemonium have given rise to recent neural network models that have also opted for feature-based approaches to object representation. One example of a recent connectionist feature-based model is Mozer's (1991) MORSEL network, which was designed to represent multiple objects—specifically, multiple words. MORSEL has several components, the most relevant being BLIRNET, which gets its name because it Builds Location Invariant Representations. BLIRNET, an object representation system, creates invariant object representations by gradually collapsing across space from one layer of representation to the next. This approach to object representation involves a convergence of information as one progresses upward in the hierarchy, as is shown in Figure 1. For example, in Layer 1 of the network, there are individual image features (e.g., oriented edge segments) in particular spatial or retinotopic locations. As one progresses to Level 2 of the network, the receptive fields of units in this layer are larger, thus allowing these units to start to code for more complex features (e.g., T junctions or L junctions) with less of a reliance on where that feature is located. This spatial collapsing continues, so that by the time information reaches the highest levels in the network the "features" responded to by individual units are relatively complex and correspond to objects or components (parts) of objects. BLIRNET also possesses the ability to learn these feature combinations, allowing the model to be sensitive

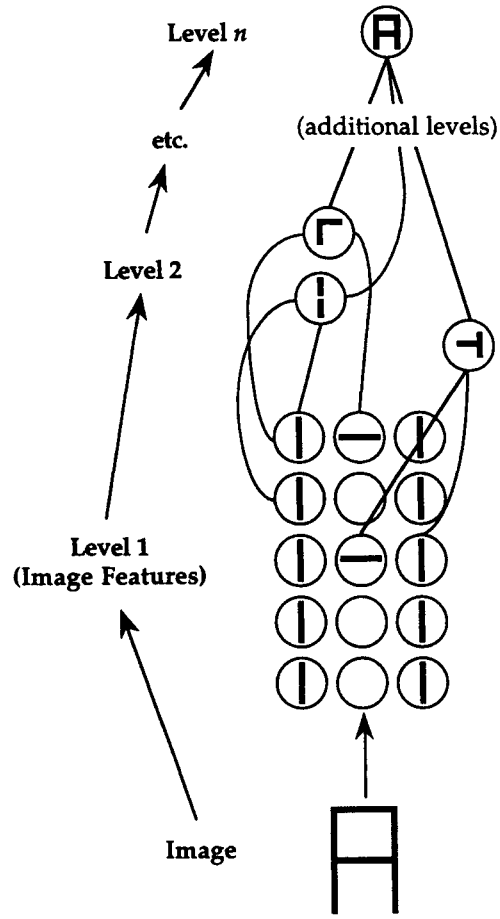

Figure 1. Example of the hierarchy in Mozer's (1991) BLIRNET, which represents objects by collapsing over spatial information as information progresses to higher levels of the hierarchy. See text for additional discussion.

to the statistical regularities in the visual environment. Unlike Pandemonium, the object representations created by BLIRNET through experience are distributed representations in which multiple units in the network code an object.

Other feature-based models have used more advanced computational techniques in order to refine the feature-based approach to object representation and to allow these systems to represent objects that are more complex than simple letters and words (for a review, see Edelman, 1997). To create object representations in which an object's features or parts are coded with respect to a reference point on the object itself (i.e., to code an object-centered representation), many feature models have relied on geometric constraints. The most relied-on constraint has been the *viewpoint consistency constraint* (see, e.g., Lowe, 1987b), in which all the features of an object are interpreted as being consistent with viewing that object from a single viewpoint. This constraint, although seemingly simple, is powerful, in that it allows object representations to remain invariant across different viewing conditions (e.g., size,

spatial position), provided that correspondence can be established between the features in the input image and the stored object representation. Furthermore, the power of this constraint allows some vision systems to establish object representations despite changes in the viewpoint or orientation of objects. For example, in Ullman's (1989, 1996) alignment approach, having a stored set of features or points at known locations on an object allows the input image to be transformed, or aligned, so that the object in that image can be determined and represented.

Although the viewpoint consistency constraint is powerful, there are limitations to feature-based models based on alignment approaches (see, e.g., Lowe, 1987a; Ullman, 1989, 1996). For example, such systems do not work well for nonrigid objects, objects such as human or animal forms, whose internal geometry can change on the basis of the movement of a part, such as crossing a leg or moving an arm across one's midline. A more serious problem for such approaches is knowing how to align the image features before object representation. Knowing the object that is present would allow the correct alignment to be performed, but object representation cannot occur prior to alignment. To overcome such difficulties, some feature-based systems have represented objects on the basis of a larger number of features, instead of relying on a small number of features, as is advocated by alignment approaches. A larger number of features allows an object to be represented as a vector in a high-dimensional feature space. This approach has the advantage that, if some small number of features missing (because of occlusion or other image variability), the object representation will not suffer greatly; the vector determined by the features will still point in the same general direction in the feature space. Some of these systems learn the feature vectors, allowing the system to make use of the statistical regularities within object classes to establish object representations (see Mel, 1996); adaptive systems (i.e., systems that learn object representations) have the added advantage of potentially suggesting learning mechanisms that biological vision systems may use in creating object representations.

As with many things, however, more is not always better. Specifically, in the case of dimensionality in object representation, using high-dimensional feature spaces raises some computational problems. As Edelman (1997) has pointed out in his review of computational approaches to recognition, learning object representations via high-dimensional feature spaces is computationally difficult. Mathematical analyses of this issue indicate that, as the feature space increases in dimensionality, the number of learning examples required to create an object representation increases exponentially. Thus, an implausible number of learning trials might be required to learn object representations. Computationally, an optimal solution to the problems raised in feature models would be to rely on a larger number of features than do alignment models but to reduce the dimensionality of the feature space, so that large numbers of training examples would not be required to acquire object representations. That is, a balance needs to be struck between the alignment approaches and the high-dimensional feature space approaches.

Edelman and his colleagues (e.g., Edelman, 1995; Edelman & Duvdevani-Bar, 1997; Edelman & Weinshall, 1991; Poggio & Edelman, 1990) have developed computational approaches with which they try to find this balance by using dimensionality reduction in representing objects. In this approach, a large number of features are sampled from the input image, resulting in a high-dimensional measurement space (Edelman & Duvdevani-Bar, 1997). The features could be, for example, the intensity at every pixel location in an image. The dimensionality of this high-dimensional feature space is then reduced by comparing the feature space representation with a fixed number of reference shapes that are stored within the system. In short, the system represents objects on the basis of the similarity between the object in the image and a set of internally stored prototypes. There exist several prototypes of individual objects, each prototype storing a different view of the object; this amounts to a few views of each individual object being stored in visual memory. Note that these prototypes are not rigid templates, because they are not tied to specific retinal locations. Under this representational scheme, novel objects and novel views of known objects can be represented by interpolating among the stored prototypes and computing the similarity between a new object or view and the stored objects and views. This prototype scheme has the advantage that it not only allows the recognition system to deal with novel objects and views but also readily permits categorization of objects (i.e., classifying an object broadly as a *car* or as an *airplane*). Many recognition systems are designed to recognize specific instances of objects (e.g., *Volkswagen* or *Boeing 727*) and do not allow for easy categorization.

Feature-based models represent a wide range of approaches to object representation. As these models have developed, the notion of a feature has changed dramatically from the image edges and junctions in Selfridge and Neisser's (1960) model to the feature spaces, based on large samplings of image data, used in current systems. The importance of learning object representations has also emerged in feature models, with feature vectors (Mel, 1996) or object prototypes (see, e.g., Edelman & Duvdevani-Bar, 1997) being acquired through experience with objects that appear at different orientations or in different views. The use of learning algorithms may be important for understanding biological object representation, which must acquire object representations through experience.

Volumetric Models

One potential limitation of feature models is that the object representations in many of these systems do not explicitly code the parts of objects and the relations among the parts (see Pinker, 1984). For example, in approaches that rely on vectors in high-dimensional feature spaces (e.g., Mel, 1996), the parts of an object contribute

to the object representation (i.e., the vector), but the representation does not code for the relative positions of the parts of an object. Failure to explicitly code the relations among the parts could cause problems for recognizing objects that have similar part configurations (e.g., a cup, which has a handle on its side, and a pail, which has a handle on top).

Volumetric models, in which the three-dimensional (3-D) structure of an object is explicitly represented, provide an alternative to feature models and overcome the problem of part representation that feature-based models possess. Volumetric models rely upon structural descriptions, which are explicit descriptions of an object's structure that specify the parts of an object and the relationship among these parts. (However, other computational accounts may also use structural descriptions.) For example, a structural description of a human might have the arms being coded to the left and right of the torso, the legs being coded below and to the left and right of the torso, and the head being centered above the torso. A structural description would not be verbal, as in the foregoing example, but would instead involve image parameters, such as the orientation of the part or a reference point on a part, that would specify the position between some part and either the object or a higher level part.

Marr and Nishihara (1978) provided one of the earliest and best-known volumetric models of object representation. Their approach involved creation of a structural description in which each part was coded relative to a reference point (or origin) that was centered on the object (i.e., the structural description was coded in an object-centered reference frame, a reference frame in which a point of reference lies on the object itself); the resulting object representation is referred to as a 3-D model. Marr and Nishihara's 3-D model was a volumetric representation, in that the parts were represented as generalized cylinders, a volume that can be created by sweeping out a circle along an axis. Some object parts, such as a human torso, could be represented by a generalized cylinder that has some diameter and some length; for example, the cylinder used to represent a torso would have a larger diameter and would be longer than a cylinder that was used to represent another part of a human, such as the arm. The parts (i.e., the generalized cylinders) could be specified precisely by parameters defining the diameter and length of the cylinder that represented the part. Generalized cylinders also permit the axis of elongation (i.e., the axis of the cylinder) of a part to be extracted, which allows the viewer to determine the overall orientation of that part.

One important contribution of Marr and Nishihara's (1978) recognition scheme is that it was one of the earliest computational accounts of object representation that made explicit the mapping from low-level perceptual representations to object-based representations. In the Marr and Nishihara approach, the 3-D model is an object representation. The inputs to the 3-D model come from the 2½-D sketch, a viewer-centered perceptual representation that codes the visible surfaces of an object or scene. The 2½-D sketch does not possess any of the invariances that object representations require to overcome variability in input images, such as spatial invariance or size invariance, which limits the usefulness of this representation for object recognition. To achieve the invariances needed for object recognition, the 3-D model was extracted from the 2½-D sketch. Although Marr was never completely successful at bootstrapping the 3-D model from the 2½-D sketch, he did demonstrate some limited recognition, using silhouette shapes (Marr & Nishihara, 1978). The system would first decompose the image of an object into likely parts, using points of extreme curvature (i.e., points in which curvature changes greatly; also see Hoffman & Richards, 1984). The major axis of elongation for each part would then be computed, and this axis would be used as the axis of elongation of a generalized cylinder that would represent that part. The collection of generalized cylinders for each of the parts of an object forms the 3-D model.

Although Marr (1982) provided many good arguments for the 3-D model, his approach to object representation has fallen out of favor because of a number of limitations, the most significant being the limitations of generalized cylinders as primitives for object representations. Generalized cylinders are sufficient for representing some objects, such as human and animal forms, but are insufficient at representing other types of objects, such as desks, trees, and telephones. As a response to the limitations of cylinders as part primitives, other computational vision systems have developed volumetric primitives that extend beyond cylinders. Other volumetric primitives have been developed in order to overcome this limitation; for example, Pentland's (1986) superquadric volumes allow other volumetric shapes (e.g., cubes) to act as parts of objects.

Perhaps one of the most influential volumetric models, particularly in cognitive psychology, is Biederman's (1987; Hummel & Biederman, 1992) recognition-by-components (RBC) account of object recognition. Biederman's RBC model proposes volumetric primitives known as geons (for *geometrical ions*). Geons are volumes that can be modeled as generalized cones, which are volumes created by sweeping out a cross-section along some axis. Generalized cones are not as narrowly defined as generalized cylinders for two reasons: (1) The shape of the cross-section can vary (i.e., the cross-section does not need to be a circle, as in a generalized cylinder), and (2) the axis can curve, expand, or contract (i.e., the axis does not need to remain straight, as in a cylinder). Thus, geons have different shapes, allowing them to represent the parts of many different objects. Examples of geons and the creation of geons are shown in Figure 2.

Geons can be viewed as the primitives for object representations, much as phonemes are the primitives for speech. The advantage of having a well-defined set of primitives is that a relatively small number of primitives (i.e., geons) and a small number of relations among the

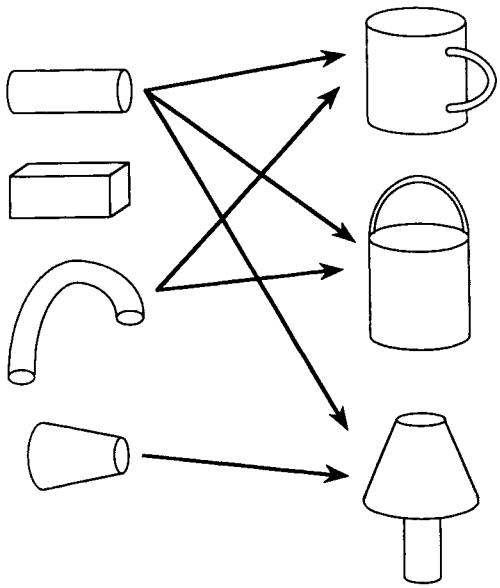

Figure 2. Examples of four geons (a cylinder, a brick, a tube, and a wedge) and how different objects can be created from a combination of different geons.

primitives permit a large number of objects to be represented. Biederman (1987) has calculated that, with only 36 geons, almost 75,000 two-geon objects could be represented (assuming different sizes and relationships between the two geons), just as the (approximately) 43 phonemes in the English language allow for the creation of all possible words.

Biederman (1987) defines geons on the basis of nonaccidental properties (Lowe, 1985). Nonaccidental properties are image properties, such as parallelism and points of cotermination, properties that are unlikely to arise in an image because of an accidental viewpoint. When present in an image, nonaccidental properties can be assumed to reflect accurately the external stimulus. Defining geons with nonaccidental properties allows geons to be viewpoint invariant; that is, an individual geon will be recognizable across most views of that geon. (Of course, there are limitations to viewpoint invariance, such as foreshortening, in which a geon is viewed parallel to one of its major axes. For example, viewing a brick end on only projects a squarish retinal image and provides no information about the volume of the brick.)

Under RBC, object representations are created by geons, which represent individual parts of an object, and the relationships among these geons. Because the parts of an object (i.e., the geons) are invariant across viewpoints, spatial position, and retinal size, the object representations will also possess these invariances. Biederman also discusses how spatial and metric information (such as viewpoint, spatial position, and size) might be coded by a separate system that is not responsible for object representation (Biederman & Cooper, 1992); this separate system would allow metric information to be recovered when needed but would not hinder recognition by requiring a specific view of an object, a specific spatial location of the object, or a specific size of the object.

Despite the advantages of volumetric models, most notably their ability to explicitly code the parts of an object, there are drawbacks to these vision systems. One significant drawback, discussed by several authors, is the computational expense of volumetric primitives (see Plaut & Farah, 1990; but see Brooks, 1981, for a working computational system that relies on volumetric primitives). Another drawback is the computational difficulty in extracting volumetric primitives from raw image data (see Edelman, 1997); although representations of volumetric primitives can be extracted from labeled line drawings (see Hummel & Biederman, 1992), human and nonhuman primate[1] visual systems work from raw, retinal images, and any adequate computational model should permit recognition from such images.

Summary

There are several computational problems that must be addressed by any object recognition system, natural or artificial. These problems include the different invariances that appear to characterize primate object representation, as well as the problem of explicitly coding the parts of an object. Each of the computational approaches just reviewed addresses these computational problems in different ways, with some models having clear advantages over other models (e.g., the advantage of volumetric models over feature models for explicit coding of parts). No model is entirely perfect, and each model's strengths and weaknesses will need to be elucidated, not only with computational considerations, but also with considerations of behavioral data and neuroscientific data.

Having addressed some of the major computational problems and models of object representation under Marr's (1982) computational level, I now turn to a consideration of the next level in Marr's framework, the algorithm level. The algorithm level addresses the representations and processes that actually solve the computational problems. The area of research that focuses on object representations and the processes that operate on these representations is cognitive psychology. The next section focuses on behavioral results from both cognitive psychology and psychophysics that have illuminated our understanding of object representation.

BEHAVIORAL STUDIES: THE ALGORITHMIC LEVEL

The computational problems concerning object representation are problems that face any system that must process objects. Furthermore, for any problem, such as spatial invariance, there are potentially several ways in which the problem could be solved. Of particular interest for cognitive neuroscientists is how biological vision systems, particularly the primate visual system, represent objects. Most of the behavioral research on object

representation has been conducted with humans and has used paradigms from cognitive psychology and psychophysics. The theoretical issues addressed by individual studies vary greatly, and my goal in this section is not to survey the entire literature but rather to characterize the operation of human object representation. There are many other behavioral studies that have addressed important issues, such as whether object recognition requires surface representations (e.g., a 2½-D sketch) or edge representations (see Biederman & Ju, 1988), that I will not discuss. Instead, I will focus on three of the most highly investigated invariances: orientation invariance, size invariance, and transformation invariance.

The focus of most studies of object recognition is on how humans perform entry level recognition. Entry level recognition refers to categorization at the *basic level* (see Biederman, 1987), in which an object is identified as a *bird* or as a *car*, as opposed to recognition at the *subordinate level*, in which an object would be identified as a more specific instance of the object, such as a *robin* or as a *Mercedes*. Basic level recognition is emphasized in cognitive studies of human object recognition, because nonprototypical subordinates may be recognized differently than basic level objects. For example, recognizing a *penguin* as such is faster than recognizing the penguin as a *bird* (Jolicoeur, Gluck, & Kosslyn, 1984). Studies of object recognition have thus tended to study the recognition of more prototypical objects or to consider atypical subordinates (e.g., *penguins*) as occurring at the basic level of categorization. The argument in the literature has been that basic level, or entry level, recognition is more characteristic of the everyday object recognition used by humans.

Orientation Invariance?

Perhaps most of the behavioral research on object recognition has centered on the issue of whether human vision is orientation dependent or orientation independent. The question is, when recognizing an object, is the object recognized across all orientations or views of the object? *Object-centered* theories (or orientation-invariant theories) answer *yes* to this question and state that an object representation contains the geometry of the object that allows the object to be recognized irrespective of the orientation of the object. Object-centered theories predict that recognition will take the same amount of time to occur across different orientations, or views, of the object, with the constraint that the parts of the object must be visible (see, e.g., Biederman, 1987; Biederman & Gerhardstein, 1993). Clearly, unique views of objects can be created, such as viewing an object parallel to the major axis, as when viewing a car directly from the front or directly from behind. Under such views, many features and parts are obscured, which may impair the representation and recognition of the object. Thus, provided that the parts and the relationships among the parts are recoverable from the image, the specific viewpoint or orientation should not influence recognition.

By contrast with object-centered theories, *viewer-centered* theories (or orientation-dependent theories) state that object representations store specific views of an object, such as storing the frontal view of a face or the three-quarters view of a car. Because object representations store specific views of objects, object recognition is dependent on the orientation of the stimulus (and the views of that object given a particular orientation). Objects that do not appear in an orientation stored in visual memory are recognized by transforming the image with a stored view (see Tarr, 1995), a transformation that might be analogous to mental rotation. Thus, the viewer-centered account predicts longer recognition times as an object is oriented away from the view coded by an object representation stored in visual memory.

Given the object-centered and the viewer-centered alternatives, one could argue for rejecting the viewer-centered theory on a priori grounds. Because an object can appear in an almost infinite number of views or orientations, for every individual object, an almost infinite number of view-specific object representations would need to be stored in visual memory (an argument often made by those in computational vision). However, this argument assumes that the transformation between a given view and a stored view is either too slow or not robust enough to permit object recognition. Instead, one could assume that the transformation process does most of the work and that only a few views of every individual object are stored; images that contain objects in novel (nonstored) views could then be transformed to a known view for purposes of recognition. Although the transformation would take some time, it could be rapid enough to permit flawless recognition.

Which theory better explains human object recognition? Early research on mental rotation (see Shepard & Cooper, 1982, for a review) seems to support the viewer-centered account: Numerous results indicated that the orientation of an object influenced the time to process that object. For example, in determining whether a letter appeared normal or as its mirror image, Shepard and Cooper found significant effects for the alignment of the letter in the frontal plane; if the letter appeared in its canonical, upright orientation, the normal/mirror image judgment was made quickly, and as the orientation deviated from upright, the judgments took longer. The increase in response times was a linear function of angular orientation. Similar results have been obtained for more complex objects that have been rotated in depth (Shepard & Metzler, 1971).

Research that followed the mental rotation studies also suggested orientation dependence in human object recognition. For example, Bartram (1974) reported results that demonstrated an effect of orientation on object naming. Bartram had subjects name photographs of objects. After an initial set of photographs, the photographs of

the objects were altered in various ways, the most relevant way being no change in the orientation of the object (identical pictures) versus a change in the orientation of the object in depth by 45° (rotated pictures). Bartram found that subjects named the identical pictures faster than they did the rotated pictures, a result inconsistent with orientation invariance in object representation. Instead, the increased time required to name an object in the *rotated picture* condition could be explained by hypothesizing an orientation-dependent object representation that was established on the basis of the initial presentation of the object. When the object appeared rotated in a later picture, a transformation was required to align the new image with the stored representation, leading to longer recognition times.

Numerous other demonstrations of orientation dependence were reported after Bartram's studies. For example, Palmer, Rosch, and Chase (1981) reported that perceivers rate some views of objects as being better than other views. Thus, not all views of an object appear to be equivalent with one another, indicating that the view of an object influences perceptibility of that object, supporting the viewer-centered theory. Also, Rock and colleagues (Rock & DiVita, 1987; Rock, DiVita, & Barbeito, 1981) demonstrated that, after seeing a novel 3-D object (similar to a bent paper clip) in a specific orientation, subjects were extraordinarily poor at recognizing that same object when it appeared either following a rotation in depth (Rock et al., 1981) or following a shift in location that resulted in a different retinal image from the original viewing (Rock & DiVita, 1987; also see Edelman & Bülthoff, 1992, for similar results using bent wire objects).

Despite these demonstrations that appear to support viewer-centered, or orientation-dependent, accounts of human object representation, Biederman and his colleagues (e.g., Biederman, 1987; Biederman & Gerhardstein, 1993, 1995; Hummel & Biederman, 1992) have pointed out several difficulties with the foregoing results and have presented evidence favoring object-centered object representation (consistent with the RBC account). Biederman argued that mental rotation rates are too slow to explain the ease and robustness of recognition across different viewpoints. Bartram's (1974) results could be explained on the basis of an occlusion of some parts of an object (Biederman & Gerhardstein, 1993). If some of the rotated pictures resulted in occluded parts, recognition would be slowed and would be dependent on the orientation of the object. Palmer et al.'s (1981) results, indicating that some views were more canonical than others, is also consistent with Biederman's RBC account, because the views that subjects report as being the best views are likely to be special views that maximize the number of visible parts of an object. Under RBC, parts (represented by geons) provide the input to object representations, so having a large number of parts visible would maximize the match between the visible image and the object representation, thus allowing the canonical view to be reported as being a particularly good view of the object. Finally, Biederman and Gerhardstein (1993) have argued that the results reported by Rock and colleagues (Rock & DiVita, 1987; Rock et al., 1981) are problematic because the objects used, the bent paper clips, lack a critical property that RBC requires for object-centered recognition: easily identifiable viewpoint-invariant parts. Rock's bent paper clips simply do not have salient parts that would permit recognition across different viewpoints.

On the basis of these shortcomings in the previous research, Biederman and Gerhardstein (1993) demonstrated that recognition could be object centered and independent of the orientation of the object. They created stimuli that satisfied three properties required by RBC: (1) easily identifiable viewpoint-invariant parts (the property violated by Rock's bent paper clips); (2) different geon structural descriptions, which are representations of the parts (geons) and the relationships among the parts, for each different object; and (3) identical geon structural descriptions across the different viewpoints of a single object. Biederman and Gerhardstein (1993) demonstrated that recognition of common, everyday objects (e.g., a flashlight) would be orientation invariant, because the three properties were met. Similarly, recognition of entirely novel stimuli was also orientation invariant, because the novel objects were created to meet the three properties. In addition, when the same novel objects were oriented so that parts visible in an initial view of the object were now occluded, whereas other parts, previously occluded, were now visible, recognition was then *dependent* on the orientation of the stimulus. This emergence of orientation dependence, or viewer-centered representation, resulted because the new view of the object violated the third property: Because some parts were occluded and other parts became visible, the geon structural description for the object was different across the two views, thereby leading to viewer-centered recognition.

Although Biederman and Gerhardstein (1993) presented some evidence favoring object-centered representation, data favoring orientation-dependent, viewer-centered representations continued to amass. Tarr and his colleagues (Tarr, 1995; Tarr & Bülthoff, 1995; Tarr & Pinker, 1989, 1990) have continued to argue for a viewer-centered account by demonstrating that recognition depends on the view of an object. For example, Tarr and Pinker (1989) used two-dimensional (2-D) stick figure stimuli, similar to those shown in Figure 3, that were similar to one another in that each object had a main ver-

Figure 3. Two-dimensional stick figure stimuli similar to those used by Tarr and colleagues (e.g., Tarr, 1995; Tarr & Pinker, 1989, 1990) to study orientation invariance in human object recognition.

tical body and arms (or bars) that extended off of the vertical body. The subjects were trained to discriminate normal objects from the mirror images of the same objects, and, in the training phase, the subjects saw the objects in specific orientations in the frontal plane (e.g., objects could appear upright, rotated 45° clockwise, and rotated 90° counterclockwise). After performing several blocks of trials, the subjects received a *surprise* block in which the objects appeared in new orientations—orientations that differed from the orientations used in the previous blocks. The results from the surprise block demonstrated that when the objects appeared in a new orientation, the subjects appeared to align the object to the closest standard orientation. For example, consider an object that had been presented and learned in a 90° counterclockwise orientation. If this object appeared rotated 135° counterclockwise in the surprise block, the subjects would rotate the image to the standard view (90° counterclockwise) in order to perform the normal/mirror image judgment. The subjects appeared to learn specific views of the objects, and, when a new view appeared, the subjects would transform (rotate) the image to bring it into alignment with a known view. On the basis of these results, Tarr argued for a viewer-centered account of object representation that specified multiple views plus transformations (Tarr, 1995; Tarr & Pinker, 1989). In his later work, Tarr (1995) reported similar results, using 3-D objects that were rotated in depth, strengthening his argument for viewer-centered object representation. Other researchers (e.g., Edelman & Bülthoff, 1992) have reported similar results that favor viewer-centered representation.

There are two important points that have been raised about Tarr's work. First, because the objects used in Tarr's (1995; Tarr & Pinker, 1989, 1990) studies were similar to one another, recognition might have required individuation at the subordinate level, not at the level of the basic category. Thus, the recognition processes in Tarr's studies might have been different from those in other studies, such as Biederman's (e.g., Biederman & Gerhardstein, 1993), that demonstrate object-centered representation. Second, because Tarr's objects were similar to one another, they violated several of the properties that Biederman and Gerhardstein (1993) proposed for orientation-invariant, object-centered representation. Specifically, Tarr's stick objects did not possess distinct geon structural descriptions; the objects used by Tarr used the same parts (brick geons) and had similar parts (e.g., vertical bodies and arms that extend off of the body). Under Biederman's (1987) RBC account, the geon structural descriptions of Tarr's objects would be highly similar, if not identical.

This critique of both Tarr's work and the other studies demonstrating viewer-centered representation (e.g., Bartram, 1974; Edelman & Bülthoff, 1992; Rock & DiVita, 1987; Rock et al., 1981) makes several assumptions about object representation that may be dubious: The critique assumes (1) that everyday recognition involves recognition at the basic level (e.g., *bird*), not recognition at the subordinate level (e.g., *robin*), and (2) that everyday objects can be differentiated from one another on the basis of part descriptions (or geon structural descriptions). Tarr and Bülthoff (1995) addressed these two critiques of the viewer-centered approach and the assumptions made by the object-centered approach.

Tarr and Bülthoff (1995) first argued that the properties required by Biederman and Gerhardstein's (1993) object-centered, viewpoint-invariant approach could only be met by a limited set of objects and viewing conditions, limiting the generality and ecological validity of object-centered approaches. Tarr and Bülthoff next pointed out that the viewer-centered account had been supported by a tremendous range of studies that used various stimuli and methodologies, which suggested that it was likely that the results favoring viewer-centered accounts were the result of idiosyncrasies of stimuli (e.g., certain views appearing more frequently than other views) or methodological problems (e.g., using a task that taps nonrecognition systems that rely on viewpoint information). Finally, Tarr and Bülthoff argued that object-centered accounts that rely on part descriptions, such as RBC, have problems with basic-level recognition. One problem is that some different objects, such as a *cow* and a *horse*, will actually have part descriptions that are similar, which could result in the objects being assigned to the same category. A related problem is that some objects that belong to the same category could actually have different part descriptions, which would prevent these objects from being assigned to a common category; for example, Figure 3 in Tarr and Bülthoff depicts three wristwatches that should be assigned to the same category (*watch*), although these watches have very different part descriptions (or geon structural descriptions), which would make such categorization difficult. Tarr and Bülthoff's reply makes it clear that viewer-centered accounts have strong empirical support and that object-centered accounts that rely on part descriptions have their own difficulties and cannot account fully for all of the data.

To date, the research on orientation invariance in human object representation provides some evidence for both object-centered, orientation-invariant representation and viewer-centered, orientation-variant representation. Although further research on this topic surely will be forthcoming, it is likely that a resolution of the two accounts will come from arguments concerning the ecological validity and computational feasibility of the different accounts (Tarr & Bülthoff, 1995). Also, there may be multiple object representation systems, some object centered and some viewer centered, that allow these two types of object representation to coexist in the primate visual system (see Ellis, Allport, Humphreys, & Collis, 1989; Farah, 1991; Tarr & Bülthoff, 1995, for multiple object recognition systems). Knowing whether or not human object representation possesses orientation invariance will be important for understanding both human vision and the computational principles that underlie human vision.

Size Invariance

Size invariance is the ability to recognize an object irrespective of the retinal image size of that object. Recognition should occur just as efficiently when the object is close to the viewer (and casts a large retinal image) as when the object is far from the viewer (and casts a smaller retinal image). If human object representation did not possess size invariance, different object representations would be required for every retinal size at which an object could appear. On a standard computational analysis, because objects can appear in an almost infinite number of retinal sizes, too many object representations would be required to permit robust recognition and to permit new object representations to be learned. However, although there are good computational and engineering reasons to favor size invariance in an object recognition system, it is unlikely that the primate visual system evolved in accordance with the ideal computational or engineering principles. Thus, whether human object representation is size invariant remains an open issue, and several empirical studies have asked whether human shape perception is dependent on the size of the retinal image or independent of the size of the retinal image.

Early studies on human size processing indicated that the size of the stimulus influences the time required to perform a shape-matching task. Bundesen and Larsen (1975; Larsen & Bundesen, 1978) presented subjects with a simple matching task in which two shapes were either the same or different; the subjects simply had to report whether the shapes matched or mismatched. The shapes were either the same size or different sizes; when the shapes were different sizes, the size difference between the two shapes was varied. For example, two shapes could differ in size by a 2:1 ratio or by a 4:1 ratio. The results indicated that the subjects were sensitive to size differences. When the two shapes were different sizes, the subjects took longer to perform the matching task than when the shapes were the same size. Furthermore, matching time increased linearly as the size discrepancy between the two shapes increased. Thus, the results suggested that the subjects mentally scaled (expanded or contracted) one of the two shapes to match the size of the other shape, a mental scaling analogous to the mental rotation of misoriented shapes (Shepard & Cooper, 1982). Similar results have been obtained for matching shapes in visual memory tasks (Jolicoeur, 1987; Larsen & Bundesen, 1978). If the subjects see a shape followed by another shape some time later, matching reaction times are again linearly related to the size discrepancy between the two shapes. These results suggest that human object representation varies with size and does *not* possess size invariance.

In contrast with these size-scaling results, research by Biederman and colleagues (Biederman & Cooper, 1992) has demonstrated size invariance in human object recognition. Biederman and Cooper (1992) developed a simple naming task in which subjects named pictures of common everyday objects, using the basic level name (e.g., *bird*). This task was adopted to overcome problems with matching tasks, such as the possibility that a matching task could be performed at an early level of representation that was spatiotopically mapped and was not size invariant. In Biederman and Cooper's (1992) task, the subjects first named objects in an initial block; in this block, half of the objects were large (approximately 6° of visual angle), and half were small (approximately 3.5° of visual angle). In a second block, the subjects saw pictures of the same objects, but now the objects were either the same or a different size as in the first block. The critical comparison was between a change in size from the first block to the second block and no change in size from the first block to the second block. The results showed no effect of size change: The subjects were just as fast to name the object when it appeared at a different size as when it appeared at the same size. Furthermore, control conditions demonstrated that these results were not due to priming a name representation. Biederman and Cooper's (1992) results are easily accommodated by an object recognition system that is insensitive to changes in retinal size and forms object representations that are based on geometric properties of the object that are invariant across size transformations.

Because Biederman and Cooper's (1992) results were obtained with a task that forces object recognition, it seems safe to conclude that recognition is size invariant. Simple matching tasks, such as those used by Bundesen and Larsen (1975; Larsen & Bundesen, 1978), and *old versus new* judgments (Jolicoeur, 1987) may not require explicit recognition and may, therefore, rely on processing mechanisms that are sensitive to retinal size and occur before size-invariant recognition processes. One issue for further research, however, is whether familiarity plays a role in size invariance: Biederman and Cooper (1992), who found evidence for size invariance, used highly familiar objects as stimuli. Since subjects would have had the opportunity to see such objects in many different sizes, the possibility remains that the visual system stores size-specific object representations (much as it might store viewpoint-specific representations). There is, however, neurophysiological evidence, reviewed later, against size-specific object representations. Thus, although there are some additional issues (e.g., the role of familiarity) that need to be investigated, the literature seems to indicate that human object representation is size invariant when recognition tasks are used.

Translation Invariance

The final major computational challenge facing any object representation system is that of translation (or spatial) invariance, the ability to recognize an object irrespective of the spatial location of the object. In a spatially invariant recognition system, recognition should occur just as efficiently when the object appears in a new spatial or retinal location as when the object appears in a spatial or retinal position in which it was previously observed. As with size invariance, if a recognition system was not translation invariant, a separate object represen-

tation would be required for every location that an object could occupy. For example, there would be several *dog* representations, one for each retinal location that an image of a dog could occupy. Because objects can appear in many retinal locations, translation invariance would make recognition easier by reducing the number of object representations required for individual objects. Specifically, one object representation that coded an object irrespective of its retinal location would suffice for every recognizable object.

As with orientation and size invariance, empirical studies have examined whether human shape processing is sensitive to the location in which an object appears. Biederman and his colleagues have presented behavioral results that support translation invariance in human vision. Biederman and Cooper (1991) again used a priming paradigm in which subjects named visually presented objects. The subjects first named objects in an initial block; in this block, half of the objects appeared in a spatial location to the left of fixation, and half appeared in a location to the right of fixation. In a second block, the subjects saw pictures of the same objects, but now the objects appeared either in the same location or in a different location than in the first block. For example, if a dog appeared to the left of fixation in the first block, it could appear either to the left of fixation (identical position condition) or to the right of fixation (different position condition) in the second block. The critical comparison was between a change in spatial position from the first block to the second block and no change in position from the first block to the second block. The results showed no effect of spatial location: The subjects were just as fast to name the object when it appeared in a different location as when it appeared at the same location in the second block. Control conditions demonstrated that these results were not due to priming a nonvisual name representation; some objects across the two blocks had the same name (e.g., *piano*) but involved different objects (e.g., a grand piano in the first block and an upright piano in the second block). If the priming from the first block to the second block had been solely nonvisual, the subjects should have been faster to recognize and name the objects that had the same name but different images, as in the piano example. However, the subjects demonstrated no such priming, indicating that the priming effects from the first block to the second block were attributable to a visual object representation, not to a nonvisual name representation.

Studies using attentional selection tasks also support spatial invariance in human object representation. Vecera and Farah (1994) used an object discrimination task (see Duncan, 1984) to determine whether visual attention could select from spatially invariant object representations. The subjects saw displays containing two objects, a box and a line. Each object had two attributes that varied; the box was either short or tall and had a gap on the left or the right side, and the line was either dotted or dashed and was tilted either left or right. The subjects were asked to report two attributes from briefly presented displays. The attributes could come from the same object (box height, box gap) or from different objects (e.g., box height, line tilt). Furthermore, the box and line appeared superimposed on one another (the *together* condition) or separated from one another (the *separate* condition). The subjects were always less accurate in reporting attributes from different objects than in reporting attributes from the same object. However, this *object effect* did not depend on the spatial position of the objects: The cost associated with shifting attention from one object to the other was *not* increased by moving the objects apart from one another, a result consistent with attentional selection from an object representation that does not code the spatial locations of the box and line.

Despite the evidence favoring translation, or spatial, invariance in human object representation, a recent study potentially calls translation invariance into question. Dill and Fahle (1998) reported results demonstrating that human object representation may not possess translation invariance. In Dill and Fahle's experiments, the subjects were asked to indicate whether two sequentially presented objects were the same or different. The objects were novel *dot clouds* formed by placing 10 dots randomly in a square region. The two objects, which were temporally separated by a 1-sec interstimulus interval, could appear in the same location or in different locations. The reaction times when the two objects were the same showed a strong effect of spatial location: The subjects were faster to report that the clouds were the same when the dot clouds appeared in the same location than when they appeared in different locations. Dill and Fahle also conducted control experiments that ruled out the possibility that the costs associated with a shift in location were due to a shift in spatial attention from one location to another. Thus, matching two dot clouds is dependent on the location in which the stimulus appears, which implies that human object representation may be *dependent* on the location of an object.

The apparent discrepancy between Biederman and Cooper's (1991) results and Dill and Fahle's (1998) results certainly raises the need for further research on spatial invariance in human vision. There are several differences between the two studies, which makes a direct comparison between them tenuous, at best. For example, the stimuli used were highly different; Biederman and Cooper (1991) used familiar everyday objects, whereas Dill and Fahle used novel dot clouds that were quite similar to one another. The experimental procedures also differed, with Biederman and Cooper (1991) employing a naming task and Dill and Fahle employing a same–different matching task. As Biederman and Cooper (1992) discussed, tasks that do not require naming, such as matching tasks, may rely on a noninvariant level of representation that occurs prior to object representations that possess invariances (including spatial invariance). Further, the dot clouds used by Dill and Fahle were all similar to one another, which may have required subjects to be more sensitive to

metric properties of the stimuli, such as the spatial relationships among the individual dots in a single dot cloud. Reliance on metric properties could have led to the clouds being coded in a representation that did not possess spatial invariance. As with size invariance, additional research will be required on the topic of translation invariance. However, the best evidence from recognition tasks (Biederman & Cooper, 1991) indicates that human recognition may occur independently of where the object falls on the retina.

Summary

The behavioral evidence for different invariances in human vision may strike some as equivocal. For each of the invariances reviewed, behavioral evidence could be found both to support and to refute the invariance. Such contradictory results, if taken alone, would certainly paint a bleak picture for the study of human object representation. However, a more optimistic picture emerges by taking other approaches into account. For example, there are several good computational reasons, discussed previously, for each of the invariances. Neurophysiological and neuropsychological mechanisms may also provide converging evidence for or against invariances studied with behavioral methods at the *algorithmic level* of Marr (1982). The studies reviewed in the algorithmic level section are likely to show that human object representation is more complex than was previously imagined and that results from behavioral studies will depend on the tasks subjects perform as well as the stimuli presented for recognition.

To provide further insights on primate object representation, I now turn to a review of the level of implementation, Marr's (1982) *hardware* level. I have broken the review of the biological implementation of object representation into two sections, the first focusing on lesion studies and neurophysiological results from nonhuman primates and the second focusing on neuropsychological studies from humans with brain damage and on recent neuroimaging studies.

LESION AND NEUROPHYSIOLOGICAL STUDIES

The two previous sections have outlined the computational problems facing object representation systems and some of the properties of the human visual system. Each level in Marr's (1982) scheme provides useful constraints for the other levels, although the levels are independent of one another. In this section and the next, I will review the *hardware level*, the implementation of object representation mechanisms in the primate visual cortex. Neurophysiology and neuropsychology can provide strong constraints for both the computational and the algorithmic levels; for example, although the behavioral results may be contradictory, as they are with size invariance or spatial invariance, a consideration of the neural mechanisms underlying object processing may aid in resolving discrepant behavioral results.

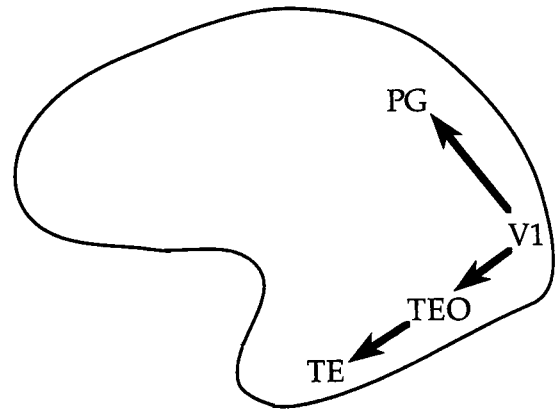

Figure 4. A lateral view of a monkey's left hemisphere showing the "what" and "where" visual pathways of Ungerleider and Mishkin (1982) as well as the subdivisions of the inferior temporal cortex.

The neural mechanisms most relevant for object representation are those extrastriate visual areas that lie along the "what" visual pathway, the pathway that extends ventrally from the occipital lobe to temporal lobe visual areas (Ungerleider & Mishkin, 1982), as is shown in Figure 4. The pioneering work of Ungerleider and Mishkin demonstrated that the ventral visual pathway was required for object-matching tasks. Monkeys with lesions to temporal lobe visual areas were impaired at matching the form of a shape across a time delay. The visual area in the temporal lobe that appears to code for object attributes and appears to represent objects is the IT cortex, which can be further subdivided. The subregions of the IT cortex that are relevant for object processing include TEO and TE; area TEO is the posterior region of the IT cortex, and area TE is the anterior region of the IT cortex (Iwai & Mishkin, 1969; Von Bonin & Bailey, 1950; see Logothetis & Sheinberg, 1996, and Tanaka, 1996, for reviews). Although the TE/TEO subdivision seems to be the convention followed most in dividing the IT cortex (see Tanaka, 1996), other subdivisions also exist, such as Felleman and Van Essen's (1991) division of the IT cortex into the posterior IT cortex (PIT), the central IT cortex (CIT), and the anterior IT cortex (AIT).

The feedforward anatomical projections into the IT cortex come from cortical visual areas V2, V3, and V4. This pathway is predominantly serial, with the information first starting in V1, which projects to V2, which then projects to V3, then to V4, then to TEO, and finally to TE. Some of the inputs to the IT cortex are *jumping* inputs that violate this serial pathway, such as direct projections to TEO from V2 that bypass V3 and V4 (Nakamura, Gattass, Desimone, & Ungerleider, 1993). This serial, feedforward pathway appears to be consistent with some computational accounts of object representation,

such as Selfridge and Neisser's (1960) Pandemonium model. This serial pathway, and the similarity to some feature-based models, may tempt some to expect "grandmother" cells (i.e., cells that respond selectively to one object, such as your grandmother's face) in the IT cortex; however, emerging evidence suggests that individual neurons in the IT cortex form part of a distributed representation of an object in which a group of neurons represents a single object (more akin to the distributed representations formed in Mozer's, 1991, BLIRNET model, discussed previously). Also, the processing between the IT cortex and other extrastriate visual areas (e.g., V4) appears to be bidirectional, with the IT cortex and its subregions sending anatomical feedback projections to earlier visual areas (see, e.g., Rockland, Saleem, & Tanaka, 1994).

Lesion Studies

What are the consequences of lesions along the "what" visual pathway in nonhuman primates? The earliest lesions specifically placed in the IT cortex (and sparing medial temporal regions) resulted in reports of visual deficits (see Dean, 1976, for an early review and Plaut & Farah, 1990, for a more recent review). For example, Mishkin and his colleagues (Mishkin, 1954, 1966; Mishkin & Pribram, 1954) reported that visual discrimination performance, in which monkeys or baboons were required to discriminate two shapes, was impaired by lesions to ventral temporal lobe areas. The hippocampus, which often was lesioned when performing lesions to the ventral temporal lobe, did not appear to result in visual discrimination impairments; lesions to the hippocampus alone left discrimination relatively intact or resulted in mild discrimination impairments (Mishkin, 1954). The impairments observed following lesions to the ventral temporal lobe exist for both postlesion retention of the discrimination (see, e.g., Mishkin, 1954) and postlesion acquisition of a visual discrimination (see, e.g., Pribram, 1954).

The discrimination deficits observed in early lesion studies of the IT and temporal cortices point to a role of these visual areas in object representation (also see Ungerleider & Mishkin, 1982). In order to discriminate two stimulus objects, presumably these objects would need to be represented for purposes of perception and comparison. Subsequent research focused on the visual information to which the IT cortex is sensitive. These studies have asked whether monkeys with IT lesions possess the invariances discussed previously—namely, spatial (translational) invariance, size invariance, and orientation invariance. Other invariances also have been investigated in IT-lesioned monkeys, such as illumination constancy (see Weiskrantz & Saunders, 1984), but these invariances will not be discussed further. The standard procedure in these studies has been to require monkeys to discriminate stimuli on some dimension, such as size, and then determine the effect of IT lesions on this discrimination.

Orientation invariance was studied in monkeys with IT lesions by Gross (1978; Holmes & Gross, 1984). Gross (1978) reported that IT-lesioned monkeys appeared to be intact in performing discriminations involving rotation of 90° or 180°. For example, IT-lesioned monkeys could discriminate an *A* tilted 90° to the left from an *A* tilted 90° to the right (a 180° difference between the two stimuli) just as well as normal (unlesioned) monkeys. The same result held for stimuli that differed in orientation by 90°: Lesioned monkeys could discriminate an upright *2* from a *2* that had been rotated 90° to the right just as well as normal monkeys. This intact orientation discrimination performance appears puzzling, considering that the same IT-lesioned monkeys were impaired relative to normals in discriminating different patterns (e.g., discriminating a * from an *o*), a discrimination that normal monkeys find easy to perform (see also Holmes & Gross, 1984, for similar results).

Gross' (1978) results appear to indicate that the IT cortex is not responsible for orientation constancy, because lesions to this visual area do not impair the ability to discriminate a shape from a rotated version of itself. The consequences of these results could be that object representation systems do not possess orientation invariance, as suggested by viewer-centered accounts of object representation. However, the results on orientation processing in the IT cortex are unlikely to distinguish object-centered and viewer-centered accounts of object representation, for the following three reasons.

The first reason that Gross' (1978) studies may not distinguish accounts of object representation is that there are difficulties with some aspects of these data, such as the stimulus changes associated with changing the orientations of objects. Large rotations—rotations of 90° or 180°—change the appearance of shapes more than do smaller rotations—rotations of 30°. For example, a *5* has a characteristic loop near the bottom of the image. If this shape is rotated 30°, that loop still remains near the bottom of the image; however, if this shape is rotated 90°, the loop now appears to the left or right, and if the *5* is rotated 180°, the loop is now on top of the image. Such image cues could be used to perform some discriminations (e.g., 90° or 180° discriminations) but would prove less useful for other discriminations (e.g., 30° discriminations). Thus, the IT cortex could code orientation-invariant object representations, and IT-lesioned monkeys could appear unimpaired because of secondary strategies (e.g., matching on image differences based on large rotations).

A second concern of Gross' (1978) data is that, although the IT-lesioned monkeys could discriminate some types of orientations, these monkeys *did* tend to perform slightly worse than normal controls (although the differences were often not statistically significant). An examination of Figure 5 in Gross (1978) reveals that, in 9 out of the 10 rotated-pattern problems, the IT-lesioned group made more errors than the normal control group, although none of these differences reached statistical significance at the .05 level. Instead of assuming that the lesioned monkeys were normal, because of nonsignificant results, there is an alternative view: If IT-lesioned monkeys were no

different from normals, then, by chance, the lesion group should have shown poorer performance than the control group only on half (50%, or 5 out of 10) of the rotated-pattern problems, not on 90% (9 out of 10) of the problems. Thus, there is some evidence for a systematic deficit in the lesioned monkeys, although this deficit may be small and difficult to detect statistically. The differences between the lesioned monkeys and the control monkeys, although not statistically significant, may reflect a real difference that is difficult to detect either because of low statistical power or because the deficit is made artificially small because of strategies that the lesioned monkeys may have used, as discussed previously.

A third difficulty with using Gross and colleagues' data to constrain theories of object representation is based on procedural differences between monkey and human studies. The procedures used in studying IT-lesioned monkeys have required monkeys to "say" (respond) that two identical patterns that differ in orientation are *different*, whereas the human behavioral studies have required subjects to say that two identical patterns that differ in orientation are the *same*. A critical question that remains to be studied in IT-lesioned monkeys is whether the monkeys could determine that a *2* and a *2* rotated 90° to the right were the same shape, a procedure closer to that used to study object processing in humans. (Note, however, that this procedure would still not distinguish object-centered and viewer-centered accounts of representation, because both of these accounts would hypothesize damage to object matching following IT lesions.) Given these concerns, other data are required to elucidate the role of area IT in orientation constancy (or orientation invariance). Such data, in the form of single-unit recordings from the IT cortex, will be considered in the following section on neurophysiological studies.

Fortunately, the role of area IT in other invariances is less complicated than its role in orientation invariance. Turning to size constancy, several studies have pointed to a clear role of temporal lobe visual areas in allowing objects to be represented irrespective of their retinal size. The earliest study on the role of the IT cortex in size constancy was by Humphrey and Weiskrantz (1969), who trained rhesus monkeys to choose the larger of two disks that were presented at different distances. To correctly perform this task, the monkeys needed to possess size constancy (or size invariance) in which the size of the disk was determined irrespective of the distance from the monkey. Size constancy is necessary in this situation because a small disk could appear to be quite close to the monkey, casting a large retinal image, whereas a large disk could cast the same retinal image as the small disk. Following training, the monkeys received either parietal lesions or IT lesions. The IT-lesioned monkeys were unable to relearn the task to criterion, although the parietal-lesioned monkeys did relearn the discrimination. Further analyses of the IT-lesioned group revealed that, when IT-lesioned monkeys made errors, they seemed to choose from one of two erroneous strategies: (1) only judging on the basis of retinal size or (2) only judging on the basis of distance. These monkeys could not combine the two pieces of visual information required for size constancy, the retinal size of an object and the distance of the object from the viewer. An inability to integrate retinal size and distance would prevent these monkeys from seeing an object as remaining constant (i.e., remaining the same object) across different viewing distances. Similar results have been reported by Ungerleider, Ganz, and Pribram (1977) and by Weiskrantz and Saunders (1984).

Finally, spatial invariance, or retinal translation, appears to be impaired in monkeys with IT lesions. Gross and colleagues (Gross & Mishkin, 1977; Seacord, Gross, & Mishkin, 1979) studied interocular transfer (transfer between the two eyes) as a special case of spatial invariance. Monkeys first were given one of three possible lesions (or no lesion) and then learned a visual discrimination in one eye. Following the acquisition of the discrimination, transfer of the discrimination to the other eye was tested. There were two main results from these studies. First, monkeys who had been given concurrent IT lesions and optic chiasm sectioning and monkeys given only IT lesions learned the initial discrimination very slowly, as compared with the normal (unlesioned) monkeys and the monkeys with only optic chiasm sections. This result confirms the role of the IT cortex in pattern discrimination and perception. Second, and more important, the monkeys who had received combined IT lesions and optic chiasm sections were unable to transfer the discrimination from the initial eye to the other eye, as compared with monkeys with only IT lesions. This second result indicates that, when visual stimuli are restricted to a single hemisphere (following the sectioning of the optic chiasm), the IT cortex is necessary for interocular transfer, because lesions to the IT cortex disrupt the ability to determine the equivalence of shapes appearing between the left and the right visual fields. This result demonstrates the importance of the IT cortex in cross-hemispheric transfer of visual information, which would be necessary in establishing object representations that remained stable as an object crossed the vertical meridian from the left visual field into the right visual field (or vice versa). Seacord et al. noted that IT mechanisms may allow for spatial invariance within a visual field.

In sum, lesion studies conducted with animals point to the central role of the IT cortex and its subregions in representing objects irrespective of sensory-based changes, such as changes in size or spatial position. These results are in accordance with some behavioral results in humans that suggest that human object representation has both size (Biederman & Cooper, 1992) and spatial (Biederman & Cooper, 1991) invariance. Although the existing data indicate that the IT cortex may not play a role in orientation constancy (see, e.g., Gross, 1978), several caveats concerning these data arose, indicating further work would be necessary on the IT cortex and the role of orientation in object representation. One difficulty with lesion studies of object representation is that they provide

little, if any, information concerning the neural representations of the IT cortex. To better understand what individual or groups of IT neurons process, we need to examine single-unit recordings from this region of extrastriate cortex.

Neurophysiological Studies

To what visual inputs would an object recognition system, such as that in the IT cortex, respond? This question has provided the motivation for most neurophysiological single-cell recordings from the IT cortex. The computational, behavioral, and lesion work reviewed so far indicates that primate object recognition appears to have certain characteristics, such as the ability to code an object's structure across stimulus-level changes. The goal of neurophysiological studies of object representation is to provide a hardware implementation that elucidates the neural representation of objects, so that we can better define terms like *code an object's structure* and so that we can understand the neural mechanisms that provide the invariances.

The earliest single-cell recordings from the IT cortex were performed by Gross and colleagues (see Gross, Bender, & Rocha-Miranda, 1969; Gross, Rocha-Miranda, & Bender, 1972), who reported that neurons in this area only responded to visual stimuli and had large receptive fields that typically included the fovea. Complex shape stimuli, such as hands, appeared to drive these neurons best. Later studies reported that the median receptive field size was approximately 25° of visual angle (Desimone & Gross, 1979; see Gross, 1992, for a review) and that most receptive fields partially extend across the vertical meridian. Stimulus selectivity seems to be constant throughout these large receptive fields, although foveally presented stimuli elicit a larger neural response than do more peripheral stimulus presentations.

Almost as soon as shape-specific neurons were identified in the IT cortex, the hypothesis was entertained that these neurons played a role in the perceptual invariances (see, e.g., Gross & Mishkin, 1977), particularly inasmuch as lesions of this visual area impaired the invariances. Schwartz and colleagues (Schwartz, Desimone, Albright, & Gross, 1983) directly tested IT cells for size, contrast, and location (spatial) invariance. The stimuli used were fourier descriptors (FDs), similar to those shown in Figure 5; FDs permit a parametric variation of shapes through changing the frequency, amplitude, and phase in a fourier expansion of a shape. This procedure for generating stimuli allows researchers to have control over the generation of shapes and the relation of one stimulus shape to another. Using FD stimuli, the tuning curve of individual IT neurons was computed, which provided the shape(s) to which individual neurons maximally responded. Once the optimal stimulus for an IT neuron was determined, the stimulus could be altered to determine whether the neural response either varied with the stimulus alterations or remained constant across stimulus alterations.

Schwartz et al.'s (1983) results indicated that the stimulus specificity observed in individual IT cells was changed little, if at all, by changing the size, contrast, or location of the preferred stimulus. For example, when the preferred stimulus was changed in size from 13^{o2} (a relatively small stimulus) of visual angle to 28^{o2} or 50^{o2} of visual angle (relatively larger stimuli), the neuron maintained its preference for the shape. This stimulus selectivity remained despite the dramatic difference in the retinal images formed by the shapes of different sizes. Thus, individual IT neurons appear able to possess size invariance by coding for an object's shape irrespective of the retinal size of that shape. Similar results were obtained for contrast changes; when the contrast of a shape was reversed (i.e., when the shape was changed from black on a white background to white on a black background), individual IT neurons retained their stimulus specificity and responded maximally to the preferred shape despite the changes in contrast. Finally, IT neurons also showed spatial invariance in that they did not change their stimulus preference when the shape was presented in a different spatial location. A neuron's preferred shape remained stable as the shape was moved 5° of visual angle into either the upper or the lower visual field and as the shape moved 5° of visual angle into either the contralateral or the ipsilateral visual field. IT neurons thus demonstrate the ability to represent a shape's structure or geometry irrespective of where the shape appears within the visual field. (Similar results have been reported for single-unit studies in which faces were used as stimuli; see Rolls & Baylis, 1986.)

Although Schwartz et al.'s (1983) data speak to many of the invariances discussed in this review, Schwartz and colleagues did not address the question of whether IT

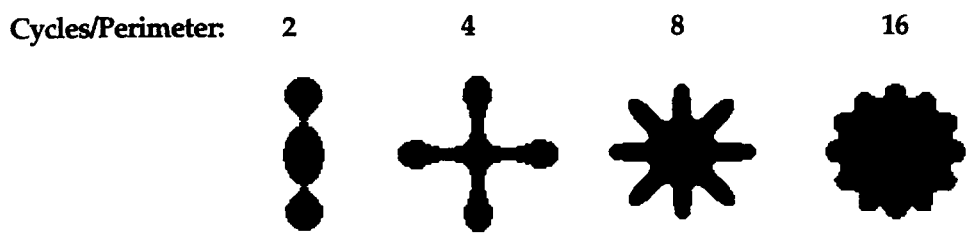

Figure 5. Examples of Fourier Descriptor stimuli used by Schwartz, Desimone, Albright, and Gross (1983) to study perceptual invariance in inferotemporal cortex neurons.

neurons possess orientation invariance. That is, when the orientation of a shape changes, would the IT neuron continue to show stimulus specificity (consistent with orientation invariance), or would the neuron cease to fire because of the orientation change (consistent with orientation variance, or viewer-centered representation)? Inspection of the FDs in Figure 5 demonstrates that a majority of these stimuli would be insufficient for studying orientation invariance, because changes in orientation would result in little, if any, change in the retinal image (the exception would be the FDs defined by frequencies of two or four, which could change their retinal images dramatically if rotated 90° for the frequency of two or 45° for the frequency of four).

However, several studies of orientation constancy in IT neurons have been reported recently by Logothetis and colleagues (see Logothetis & Pauls, 1995; Logothetis, Pauls, Bülthoff, & Poggio, 1994; Logothetis, Pauls, & Poggio, 1995; Logothetis & Sheinberg, 1996). Logothetis and colleagues trained monkeys to recognize various novel objects at a large number of orientations, so that the monkeys' behavioral performance was independent of the specific orientation (or view) of the object. Following this training (which is quite extensive and requires the monkeys to perform at 95% recognition rates or higher), individual neurons in the IT cortex were recorded. During this recording session, monkeys were first shown a target object in a *learning phase*; the target object was one of the previously learned objects presented at some orientation. Next, the monkeys were shown a series of up to 10 objects in a *test phase*; the objects had to be categorized as either a *target* (i.e., matching the previously observed target) or a *distractor* (i.e., being an object that had not been previously learned). The targets in the test phase were presented at various orientations relative to the target shown in the initial learning phase. Using this procedure, Logothetis and colleagues could determine whether individual IT neurons would respond to a target object independent of the object's orientation (or the view of that object) or whether IT neurons would code for a particular object in a particular orientation (i.e., a specific view of the object).

Several theoretically important results arose from Logothetis et al.'s studies. First, Logothetis et al. (1995) reported that individual cells in the IT cortex responded selectively to the objects that had been learned. This result is important, because the objects were initially novel to the monkeys and did not correspond to any object that the monkeys would have been familiar with prior to training (e.g., bent paper clip objects, similar to those used by Edelman & Bülthoff, 1992), indicating that IT neurons can change with experience and with the recognition requirements that face the animal (see also Miyashita, 1993). Second, although a monkey's behavioral recognition was orientation independent, individual cells in the IT cortex changed their responses with changes in object orientation. That is, the neurons in the IT cortex appeared *not* to possess orientation invariance. Instead, individual IT neurons appeared to code specific views of objects; for example, one neuron might code the left profile of an object (i.e., a view of the object rotated 90° from the frontal view), whereas another neuron might code for the right profile of the same object. (Similar results have been obtained for single-unit studies investigating face recognition; see Perrett et al., 1985.) Third, there was a small number of neurons that appeared to code objects irrespective of orientation, consistent with an object-centered, orientation-independent representation. However, there were too few of these neurons in Logothetis et al.'s (1995) sample to understand their potential role in object representation.

The majority of Logothetis et al.'s single-unit responses are consistent with viewer-centered accounts of object representation that have arisen from cognitive psychological studies with human subjects (see, e.g., Edelman & Bülthoff, 1992; Tarr, 1995; Tarr & Bülthoff, 1995) in which orientation invariance is based on multiple viewer-centered representations. The individual IT neurons may form the neural basis of the individual viewer-centered representations. An object-based, orientation-independent representation may emerge from several of these viewer-centered neurons being simultaneously active in a distributed representation (or population code). If this conclusion is supported through further work, it is important to note that individual IT neurons would *not* form an object-centered representation but rather that an object-centered representation would emerge from multiple viewer-centered representations (e.g., see Tarr, 1995). The distinction between the object-centered and multiple viewer-centered representations is subtle, but the distinction is important for theories of object representation and for understanding the neural implementation of object representations. The behavioral consequences of these two types of representation may be the same at Marr's (1982) algorithmic level, but the hardware implementation would differ.

In addition to understanding the perceptual invariances by looking at IT neurons, single-unit recordings from this visual area can potentially inform theories of object representation by suggesting what features these neurons represent. Recall that computational accounts of object representation have hypothesized different representation schemes, such as object representations based on image features such as line segments (i.e., feature-based models) or object representations based on volumetric primitives. Do IT neurons represent objects using features or primitives that can be mapped onto any of the computational accounts discussed earlier (e.g., volumetric primitives)? As with any active research area, the critical experiments either have not been performed or would be technically difficult to perform to link IT neurons to computational accounts of object representation. How-

ever, there has been active study on the representational scheme that might be used by IT neurons, and the findings of these studies may guide the development of biologically plausible computational object representation systems.

The recent work of Tanaka and his colleagues (Kobatake & Tanaka, 1994; Tanaka, Saito, Fukada, & Moriya, 1991; Wang, Tanaka, & Tanifuji, 1996; see Tanaka, 1992, 1996, 1997, for reviews) has greatly illuminated the representations that the IT cortex creates to represent and distinguish objects. Tanaka and colleagues have performed extensive single-cell recordings from both the posterior and the anterior IT cortex while using a reductive technique to determine the stimulus selectivity of the cell. This reductive technique involved finding a visual object to which the neuron responded vigorously. When such an object was found, the image was simplified as the cell's responses were recorded. For example, if the cell initially responded to an apple with a stem, the neuron would be tested with round shape, with red patches that had different shapes, with different shapes that had stems protruding from them, and so forth. Tanaka and colleagues would continue to reduce the stimulus until the stimulus could not be simplified further without abolishing responses from the cell. The simplest stimulus that sufficiently activated a cell was referred to as the critical feature (see Tanaka, 1992, for a review of this procedure).

The critical stimulus features to which IT neurons were sensitive depended on the anatomical location of the neuron. Neurons in the posterior region of the IT cortex (PIT or TEO) were more likely to be classified as *primary cells* that had critical features corresponding to relatively simple visual input. The primary cells tended to respond to oriented bars, colors, or oriented color patches. Of the cells recorded in PIT, 72% were classified as having primary critical features (Tanaka et al., 1991). By contrast with PIT, neurons in the anterior IT cortex (AIT or TE) were less likely to be classified as primary cells (only 12% received this classification). Instead, AIT neurons were more likely to be classified as *elaborate cells*, cells having critical features that required a specific shape or a combination of a specific shape and either color or texture. Also, elaborate cells could not be described as responding to some simple feature, such as an oriented bar. In the sample of AIT neurons studied by Tanaka et al., 45% were classified as elaborate; only 9% of the PIT neurons were classified as elaborate cells.

Tanaka et al.'s (1991) results suggest that there may exist a complexity continuum in the IT cortex, with more posterior regions coding for simpler features than do more anterior regions. The IT cortex is also organized in a vertical manner, in that neurons that code similar features are more likely to be contained within the same cortical column and neurons that code different features are likely to be contained in different columns (Fujita, Tanaka, Ito, & Cheng, 1992; Wang et al., 1996). One important point, however, is that the complex features responded to by AIT neurons were not specific enough to classify or recognize a single object (i.e., they were not "grandmother" cells). Thus, unlike face-specific cells found in some regions of the IT cortex, it is likely that object representation that occurs in AIT (and possibly PIT) relies on a distributed representation involving simultaneous activation of features across different cortical columns. Further studies of IT neurons will need to be performed to understand better the anatomical mechanisms that allow distributed object representations in the IT cortex. Other neuroanatomical studies of the visual cortex have revealed long-distance axonal connections that may link different cortical columns (see Rockland & Lund, 1982, for an early description of the relevant anatomy), and these connections could link the different feature columns in the IT cortex reported by Tanaka and colleagues.

What implications does Tanaka's work have for computational and behavioral studies of object representation? The most direct implications are for computational models and the representations assumed by individual models. For example, Tanaka's single-unit recording data may seem problematic for systems that use volumetric primitives for object representation. The reductive technique used to identify critical features typically begins with the presentation of 3-D objects; then, 2-D images are used to further reduce the critical feature. The important finding is that 2-D features (e.g., disks, wedges) may be the critical features of these neurons, which could be a problematic finding for volumetric representations that require 3-D features such as cylinders or bricks. However, theorists endorsing volumetric features could argue that some 2-D features would activate neurons that were ideally tuned to volumetric primitives (generalized cylinders or geons). Specific single-unit recordings may be required to test between volumetric primitives and other primitives, such as elaborate features that could be consistent with feature-based models.

Summary

The visual areas in the IT cortex clearly play a significant role in object representation. Of course, this does not exclude the importance of other visual areas, as other areas will be required to provide the IT cortex with adequate visual input. Also, there may be redundant coding of object features or object representations in other brain regions (neurons in the frontal lobe that allow objects to be held in working or reference memory could provide a redundant coding of objects; see, e.g., Rao, Rainer, & Miller, 1997). The IT cortex appears to code for stimuli despite changes in size or retinal position, but the most elegant data to date indicate that these neurons may not represent an object independent of its orientation (see Logothetis & Sheinberg, 1996). Instead, IT neurons may code for particular views of objects, thus contributing to viewer-centered object representations. Also, neurons in the IT cortex, particularly AIT, respond to complex features and may contribute to a distributed object representation.

One issue that remains is the precise role of IT neurons in object representation. Tanaka has demonstrated that IT

neurons may respond to certain features of objects. However, because the features of everyday objects are so complex, it may prove impossible to discern the unique feature coded by an individual IT neuron, because not all feature combinations can be tested in a recording session (Young, 1995). (Faces are an exception, however, because they share common parts and features, potentially making it easier to understand face representations in the IT cortex; also see Young, 1995.) A more useful approach seems to be to integrate single-unit recordings with computational accounts that use distributed representations. Such an integration places less emphasis on the specific feature coded by an IT neuron and stresses the participation of a single neuron in a distributed representation of an object.

The large receptive fields of IT neurons have been heralded as a mechanism for ignoring scale or position differences and providing the perceptual invariances. In effect, individual IT neurons "see" large portions of the visual field and respond to their critical feature anywhere within their large view of the visual field. The difficulty with this intuitive argument is that large receptive fields may have several computational properties beyond providing perceptual invariances. An excellent example comes from Hinton, McClelland, and Rumelhart (1986), who note that large receptive fields, paradoxically, allow for *better* stimulus localization than do smaller receptive fields. This computational property of large receptive fields seems at odds with lesion studies of the IT cortex in which monkeys with IT lesions continue to localize objects accurately (Ungerleider & Mishkin, 1982). Furthermore, many computational models rely on large receptive fields for invariant object representation (see, e.g., Mozer, 1991). The differences between localization and invariant perception may not only lie with the size of the receptive field but also with inputs and critical features of the neuron or unit that has the large receptive field.

Lesion studies and single-cell recordings have provided many important data concerning the neural basis of object representation, and these data can be integrated with computational approaches to representation. However, the lesion data, and especially the single-cell data, do not always make clear predictions concerning behavioral performance in humans, such as performance in the studies reviewed in the previous section. Furthermore, there exist problems for comparing human object recognition with nonhuman primate recognition. First, the presence of language systems in the human brain may alter the object representations themselves: The object representations required to name an object (as humans do) may be different from those representations needed to act on an object without naming it (as monkeys do). Second, the neuroanatomy of the ventral processing pathway obviously differs between humans and nonhuman primates because of differences in brain size. The regions in the macaque brain that participate in object representation may receive slightly different inputs than do the homologous regions in humans. Third, the experimental object recognition tasks performed by humans are often substantially different than those tasks used to study recognition in monkeys. Monkeys typically need extensive training to perform recognition tasks. This training may alter object representations or may involve the development of strategies to perform the task, strategies that would not be needed in everyday object recognition.

These concerns are not fatal, and they are not intended to belittle the importance of lesion and recording studies. What these concerns suggest is that it will be important to have converging research results from human subject populations. This evidence has typically come in the form of neuropsychological and neuroimaging studies, to which I now turn.

NEUROPSYCHOLOGICAL AND NEUROIMAGING STUDIES

As with lesion and single-cell recording studies, neuropsychological deficits and changes in regional cerebral blood flow can provide implementation constraints on both computational and behavioral approaches to object representation. Animal studies and human studies are also complementary. For example, although recording data may be useful in influencing the representations used by computational models, the same data may not be as readily useful for behavioral studies, indicating the need for neuropsychological and neuroimaging research. Conversely, neuropsychological and neuroimaging data may be more useful in interpreting behavioral studies and not as useful as single-cell recording data for computational approaches, necessitating single-unit recording data and the results from lesion studies. Thus, any cognitive neuroscience approach to object representation will need to consider relevant results from all of these methods to provide a more accurate picture of the *hardware implementation* level in object representation.

In human neuropsychology and neuroimaging studies, the neural mechanisms that have been found to be most relevant for object representation are, as in macaques, again the extrastriate visual areas that lie along the "what" visual pathway. There are, however, anatomical differences between the human "what" pathway and the corresponding pathway in monkeys. For example, human neuropsychology and neuroimaging studies show that the critical areas for object representation may be more posterior than the critical areas in monkeys. This difference was supported by recent comparisons between the human and the macaque visual cortex, using cortical flat maps in which the cortical surface is flattened for neuroanatomical visualization. Recent comparisons between the human and the monkey visual cortex, using cortical flat maps, have revealed differences in size and specific placement of visual areas (Van Essen & Drury, 1997). Study of the neural mechanisms of object representation will, therefore, require integration across such anatomical differences. The visual areas of relevance in the human visual cortex include the regions around the occipito-

temporal boundary, which, when damaged, appear to cause deficits in object recognition (see Farah, 1990, for a review). Other areas involved in human object processing include the fusiform gyrus, which has shown increases in regional cerebral blood flow (rCBF) during some object-processing tasks (see Haxby et al., 1994; Kanwisher, Woods, Iacoboni, & Mazziotta, 1997; McCarthy, Puce, Gore, & Allison, 1997).

Neuropsychology of Object Recognition

Damage to occipitotemporal visual areas often results in the syndrome of visual agnosia, an inability to recognize familiar, everyday objects. Neuropsychological studies of the agnosias have uncovered several different subtypes of this syndrome (see Farah, 1990, and Humphreys & Riddoch, 1987, for reviews), indicating that several cortical visual areas may be important in representing and recognizing objects. The earliest subdivision of the agnosias came from Lissauer (1890/1988), who described two forms of agnosia: apperceptive agnosia and associative agnosia. Apperceptive agnosia is an inability to recognize objects because of damage to early visual cortices (occipital lobe visual areas), which presumably results in problems with elementary perceptual processes. Patients with apperceptive agnosia often cannot copy pictures of visually presented objects, indicating problematic perceptual representations. Associative agnosia, by contrast, is an inability to recognize objects because of damage to later visual cortices (occipitotemporal regions, possibly corresponding to the IT cortex). Because of the apparent damage to temporal lobe visual areas, the cases of associative agnosia are more relevant to the present discussion than are the cases of apperceptive visual agnosia.

Patients with associative agnosia often have relatively intact lower level perceptual processes, such as visual acuity and spatial frequency perception (but see Bay, 1953; Bender & Feldman, 1972, for contrasting views). Associative agnosics appear to have intact picture copying, attesting to their intact perceptual processes, although these same patients are typically unable to name the object that they have copied. In addition, knowledge about objects also appears to be intact in associative agnosics; for example, these patients can often recognize objects when presented in other modalities, such as touch (see Farah, 1990, for a review).

Despite their apparently intact perceptual processes and knowledge of objects, associative agnosics show profound object recognition impairments as measured by naming visually presented objects. For example, Farah, Hammond, Levine, & Calvanio's (1988) patient L.H. was able to recognize only 73% of the 260 simple line drawings that he was shown; Wapner, Judd, and Gardner's (1978) patient was substantially more impaired, recognizing less than 25% of the objects with which he was tested. Associative agnosics also appear to use some shape information in attempting to infer what an object is. For example, Rubens and Benson's (1971) patient "often misread *K* as *R* and *L* as *T*" (p. 309), and Ratcliff and Newcomb's (1982) patient made visual mistakes in attempting to recognize objects, such as calling an anchor an umbrella, presumably because of the width of the anchor at the bottom.

If one assumes intact perceptual representations in associative agnosia, the recognition impairments observed in these patients might not involve visual object representation but semantic or linguistic processing. That is, these patients may have "a normal percept stripped of its meaning," as Teuber (1968) described the syndrome. However, many recent analyses of associative agnosia have challenged the assumption of normal low-level perception in these patients. It has become increasingly clear that these patients have subtle problems in perceptual processes. These perceptual problems constrain theories of associative agnosia and the functioning of temporal lobe visual areas in humans. Associative agnosia may indeed involve problems in visual object representation, and some of these object representation problems may involve early perceptual processes that are required for normal object representation.

The perceptual disturbances observed in associative agnosia are numerous and subtle. Although these patients can copy visually presented stimuli accurately, the procedure by which they copy appears to be abnormal. The copying is slow and slavish (see Farah, 1990). Associative agnosics might draw only one or two lines at a time or might lose their place during copying, resulting in repeated copying of parts of the stimulus (see Wapner et al., 1978, for an example). These patients are also quite sensitive to the perceptual quality of visual objects, having difficulties when pictures of objects are impoverished. The effects of visual quality can often be observed in comparing recognition between line drawings and photographs; line drawings are more impoverished than photographs because of the lack of surface detail, shadows, and so forth, and often associative agnosics have greater difficulties recognizing objects depicted in line drawings than those depicted in photographs. Additionally, some associative agnosics also fail to appreciate the differences between possible and impossible figures (Ratcliff & Newcombe, 1982), indicating that the patient's representation of overall structure, or structural description, of the shape may be abnormal or unanalyzed. All of these results indicate that low-level perceptual processes may not be as intact as some have thought.

If one assumes that object representation processes have been damaged in these patients, how do these patients perform when the same object varies in size, position, or orientation? That is, do associative agnosics show impairments in the visual constancies, similar to monkeys with IT lesions? Unfortunately there has not been a substantial amount of research on this question, so direct comparison between human neuropsychology and lesion studies in monkeys is difficult. In the cases where the question has been addressed, the results appear equivocal. On the one hand, some associative agnosics do appear to be unable to represent objects across retinal variability,

indicating a failure of the object invariances. For example, Ratcliff and Newcombe (1982) demonstrated that an associative agnosic could not match two different views of the same object. If Ratcliff and Newcomb's patient had lost orientation-invariant object representations or had lost the ability to extract an object's critical features from a display, matching across viewpoints would be impaired. This was exactly what the patient demonstrated. On the other hand, however, other agnosic patients do not seem to be as dramatically impaired, if at all impaired, on matching objects across orientations (although these patients may not be best classified as associative agnosics but as *integrative agnosics*; see Humphreys & Riddoch, 1984, patient H.J.A., for an example).

Although the variability in neuropsychological results may appear to raise more problems than solutions for object representation, one of the major contributions of neuropsychology has been to emphasize the complex nature of object representation. The representation systems discussed previously in this paper (e.g., feature models and volumetric models) have tended to involve a single recognition mechanism—namely, recognition by decomposing an object into its parts and first recognizing the parts before recognizing the entire object (this is particularly characteristic of Marr's, 1982, and Biederman's, 1987, accounts). Neuropsychological approaches to object representation suggest that there may be multiple representation systems.

One of the earliest observations that hinted at multiple recognition systems came from studies of patients with damage to the right parietal lobe (Humphreys & Riddoch, 1984; Warrington, 1985). Patients with this damage have difficulties in matching objects in which one of the objects has been misoriented to foreshorten the major axis of the object (e.g., to view a blender from directly above, parallel to the major axis of the object; Humphreys & Riddoch, 1984). These patients also have difficulties matching objects when the lighting direction differs in the stimuli (e.g., the same object lit from above in one photograph and lit from the side in another photograph; see Warrington, 1985). On the basis of these observations, Humphreys and Riddoch (1984) suggested that there may be two routes for arriving at object constancy. One route involves computing a structural description relative to a frame of reference, and the other route involves processing distinctive local features of objects. Recent reports appear to support these two routes by demonstrating that orientation information (and possibly frame of reference information) is computed separately from object information and vice versa (Turnbull, 1997): Some patients can recognize objects but do not know the correct orientation of a visual object, whereas other patients cannot recognize objects but do know the correct orientation of a visual object (Turnbull, 1997).

A recent meta-analysis by Farah (1990, 1991) of the cases of associative agnosia also indicates that there may be multiple object recognition systems in the human brain. Farah examined the types of objects that different associative agnosics could and could not recognize. For her analysis, Farah grouped objects into three broad classes: words, common objects, and faces. Impairments in recognizing one of these classes of objects often leaves recognition of other types of objects unimpaired. For example, associative agnosics who cannot recognize common objects often have little difficulty reading words or recognizing faces. Furthermore, patients who cannot recognize (i.e., read) words, a syndrome called alexia, often show no signs of object recognition impairments. Similar results are found for face recognition. Patients who cannot recognize faces, a syndrome referred to as prosopagnosia, often can recognize other common objects. Those prosopagnosics who show some object recognition impairments for nonface stimuli (e.g., animal faces or buildings) typically are preserved in recognizing other common objects. These dissociations among recognition impairments for different stimuli suggest that there may be multiple recognition systems in the human brain, a stark contrast to computational approaches that tend to endorse a single recognition mechanism (see, e.g., Biederman, 1987; Marr, 1982). But how many recognition systems?

The patterns of dissociations discussed in the previous paragraph may indicate that there are three separate recognition systems that can be isolated from one another: a recognition system for words, a second system for common objects, and a third system for faces. Not all of these dissociations are observed in patients, however, indicating that humans may not possess three independent recognition systems. On the basis of her analysis of case studies, Farah (1990, 1991) concluded that the dissociations observed are more easily explained by postulating two object recognition systems. Farah's two-system account is depicted in Figure 6. One recognition system is hypothesized to involve part decomposition, in which objects are recognized by first breaking the object into its parts (part decomposition), followed by recognition of the parts, and finally by recognition of the object (similar to Marr's, 1982, account). Recognition by this system involves representing multiple parts. The other recognition system is hypothesized to involve no part decomposition and instead involves representing complex wholes. Recognition by this system is not mediated by prior recognition of the parts; that is, one does not need to recognize the parts prior to recognizing the whole object.

Under Farah's (1990, 1991) scheme, word recognition is thought to involve representing multiple parts (i.e., the letters), and face recognition is thought to involve recognition of the complex whole. Recognition of other common objects, objects that are neither words nor faces, relies on some mix of these two systems. Some objects, such as animal faces, may be recognized through strong involvement of the *complex whole* system and less involvement by the *part decomposition* system. Other objects may rely on the part decomposition system more than on the complex whole system. This analysis explains why

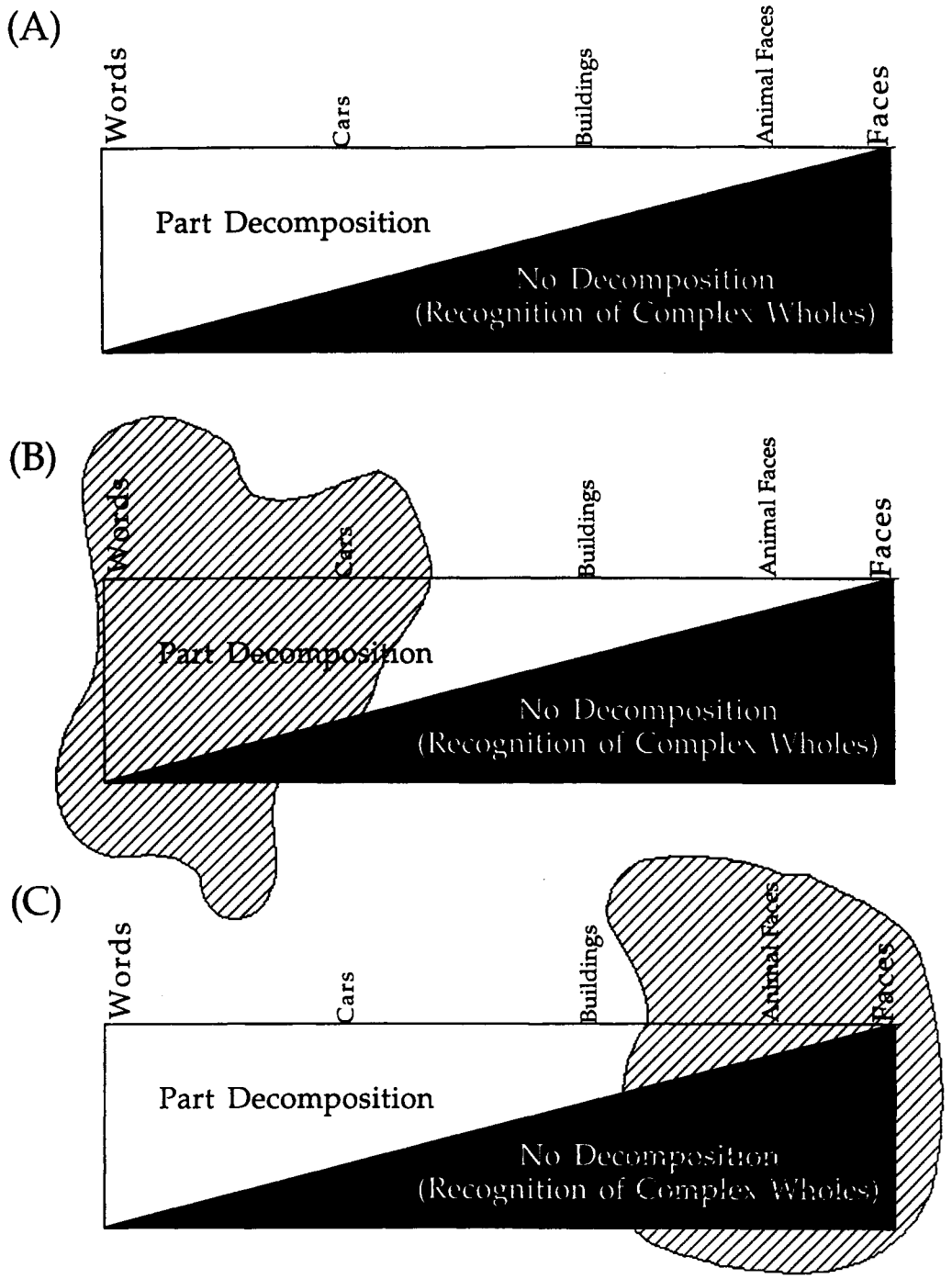

Figure 6. (A) A depiction of Farah's (1991) two-process account of object representation. Recognition of printed words taps most heavily the part decomposition system and face recognition taps most heavily the nondecomposition system. Other objects, such as buildings, cars, or animal faces, would tap the two processes by different amounts. Note that the locations of the common objects in this figure are hypothetical and for illustration only. (B) Hypothesized damage that would occur in alexia, in which visual word recognition is impaired; note that recognition of some other objects (cars in this depiction) also would be impaired. (C) Hypothesized damage that would occur in prosopagnosia, in which face recognition is impaired. Recognition of some other objects (animal faces in this depiction) might also be impaired in prosopagnosia.

some dissociations in recognition do not seem to appear frequently, if at all, such as impairments in recognizing both faces and words with intact recognition of common objects.

The possibility that multiple object recognition systems exist contrasts with what early visual theorists, such as Marr (1982), hypothesized; most theorists in computational vision (see Edelman, 1997, for a review) have tended to emphasize one recognition system that relies on part decomposition (see Farah, 1991). Indeed, part decomposition is viewed as being so important for recognition that computational vision systems that do not explicitly represent parts have been discussed as being flawed (Edelman, 1997). One challenge that will remain for multiple-systems approaches based on neuropsychology, such as Farah's (1990, 1991), will be to integrate the multiple recognition systems with both computational and behavioral data. For example, are Farah's two recognition processes each capable of computing the invariances, and are the invariances computed in the same manner by both pathways? To be consistent with behavioral and animal studies, both pathways would need to possess size and spatial invariance. Also, the debate over whether object representations contain orientation invariance may find a resolution in such a framework. One could hypothesize (1) that the part decomposition system may be orientation independent, assuming that the critical decomposed parts are visible from most orientations of the object (see, e.g., Biederman & Gerhardstein, 1993), and (2) that the complex whole system does not exhibit orientation independence, on the basis of results that demonstrate that face recognition is sensitive to orientation (see, e.g., Yin, 1969). Tarr and Bülthoff (1995) have proposed a multiple-process model similar to this analysis. Future work will need to integrate the multiple-process models that have been proposed from neuropsychology (Farah, 1990, 1991) and cognitive psychology (Tarr & Bülthoff, 1995).

Despite the advantage that neuropsychological studies have in suggesting multiple object recognition processes in the extrastriate visual cortex, research with patient populations has potential shortcomings. First, lesions that result from strokes do not obey neuroanatomical boundaries and are likely to damage multiple visual areas. Second, patients who exhibit any of the visual agnosias are very rare, which makes progress in neuropsychology slow. Third, patients are not often tested on the same recognition tasks, compounding the difficulties of comparing different patients who may have the same recognition impairments.

These possible shortcomings of neuropsychological research highlight the importance of converging methods for the study of object recognition in the human visual system. Recent advances in neuroimaging methodologies, including PET and functional MRI (fMRI), have provided another technique that has the potential to supplement the research with neuropsychological patients. For example, neuroimaging methods permit researchers to study groups of subjects and to restrict regions of interest to specific neuroanatomical areas. Although neuroimaging studies are not without their own flaws or shortcomings, imaging studies recently have contributed to our understanding of the neural mechanisms of human object recognition.

Neuroimaging Studies of Object Recognition

With the advent of PET and fMRI neuroimaging techniques, many recent attempts have been made to map both the structure and function of the human visual cortex on the basis of regional changes in cerebral blood flow (see DeYoe et al., 1996; Sereno et al., 1995). Early mapping studies demonstrated those cortical visual areas that are retinotopically mapped, such as V1 and V2, as well as visual regions that are less retinotopically mapped, if at all, such as area MT. Most relevant for the present review are those studies that have examined regional cerebral blood flow (rCBF) changes associated with object processing.

Although early studies of rCBF changes during cognitive activity revealed rCBF that could be associated with object recognition (see, e.g., Roland & Friberg, 1985), it was only the use of more sophisticated methods that allowed blood flow changes to be isolated to specific cortical visual areas. Some of the first PET studies of cognitive processes focused on reading visually presented words (e.g., Petersen, Fox, Posner, Mintun, & Raichle, 1988; Petersen, Fox, Snyder, & Raichle, 1990; see Posner & Petersen, 1990, for a review). These studies provided some evidence for the role of extrastriate areas in recognizing words, which, more generally, are a specific class of highly familiar objects. Petersen and colleagues (1990) asked subjects to passively view words (e.g., *RAZOR*), nonword consonant strings (e.g., *RGZMP*), or false font strings (e.g., ‡¬†Î◊√). Reasoning that all of the low-level visual processes involved in shape processing would be similar among words, consonant strings, and false fonts, Petersen et al. (1990) used a subtractive technique to isolate the blood flow that was specific to the recognition of visually presented words. Passive viewing of words activated a region in the left ventral occipital lobe, a region termed the *visual word form area* by Petersen and colleagues (see, also, Posner & Petersen, 1990). This visual word form area presumably corresponds to those extrastriate visual regions involved in word recognition, because this area shows increased blood flow to words but not to nonword consonant strings or false font strings (also see Puce, Allison, Asgari, Gore, & McCarthy, 1996, for a study contrasting word recognition with both face recognition and texture processing).

These early PET studies of reading provided important results concerning the representation of specific objects in the extrastriate visual cortex. However, word recognition poses several problems for understanding object recognition in generai. First, words, unlike most common everyday objects, are inherently 2-D, which limits the generalizability of the early PET studies. Second, words

are part of a larger language system and can be analyzed at different levels, such as the phonology (i.e., sound) of a word or the semantics (i.e., meaning) of a word. Third, words are also highly familiar, and such high degrees of familiarity may allow word recognition processes to become specialized and separate from general object recognition.

To overcome these limitations of word recognition studies, other investigators have studied recognition of common objects. Sergent and colleagues (Sergent, Ohta, & MacDonald, 1992) investigated both face and object recognition during PET scanning procedures. In Sergent et al.'s study, the subjects viewed line drawings of common objects, half of which depicted living objects (e.g., a horse) and half of which depicted nonliving objects (e.g., a desk); the subjects made living/nonliving judgments following each stimulus presentation. As a control task, Sergent et al. presented the subjects with sine wave gratings that varied in both spatial frequency and in orientation; the subjects were asked to determine the orientation (horizontal vs. vertical) of each grating stimulus. Blood flow changes to the grating task were subtracted from blood flow changes in the object judgment task.

Sergent et al. (1992) reported that the object minus grating subtraction revealed increases in blood flow in the left hemisphere. The statistically most reliable blood flow increases occurred in the lateral occipitotemporal areas (Broadmann's areas 19 and 20), in the left fusiform gyrus (area 37), and in the middle temporal gyrus (area 21), although this last area of activation was also found in face-processing tasks. Although the fusiform gyrus was also activated in face-processing tasks, this activation was bilateral; in the object decision task, the fusiform activation was restricted to the left hemisphere. Sergent et al. interpreted these results as demonstrating the importance of the left posterior hemisphere, particularly the lateral occipitotemporal areas, for the recognition of common objects.

More recently, Malach et al. (1995) used fMRI to identify a region in the lateral-posterior occipital lobe, termed the lateral occipital (LO) complex, that responded selectively to objects when compared with texture images. In their studies, the objects and the textures were matched on low-level image measures (Fourier power spectrum), although another set of texture images was not matched with the objects on low-level image measures. Malach et al. demonstrated that the LO complex was activated by object images but not by texture images. Furthermore, the LO complex did not appear to distinguish different classes of objects, such as common objects, faces, or sculptures. Malach and colleagues also conducted several control conditions to exclude possible artifacts that could have selectively activated the LO complex. In one study, the subjects were asked to make active scanning eye movements in one block but to hold fixation constant in another block. No differences in LO activation were found between the scanning and fixation blocks, suggesting that it was unlikely that the LO activation had been caused by scanning differences between object images and texture images. Malach and colleagues also demonstrated that activation of the LO complex was independent of low-level image activation, by (1) filtering object images, (2) adding visual noise to the object images, and (3) changing the size of object images. None of these manipulations eliminated the activation of the LO complex, provided that the object was still perceptible. That is, as long as the object was visible, albeit degraded, the LO complex showed activation. Malach et al. interpreted these results as demonstrating the importance of the LO complex in object detection. Furthermore, this region does not seem to play a role in the semantic analysis of objects, because the LO complex shows activation for unfamiliar, abstract objects such as sculptures; the LO complex also shows no activation differences between common objects and faces.

The studies by Sergent et al. (1992) and Malach et al. (1995) provided some of the important first steps in using neuroimaging to study object recognition, by using appropriate control conditions and subtractions. However, although both studies attempted to rule out low-level stimulus differences (e.g., image intensity, spatial frequency) between object and nonobject stimuli, there may be other differences in these stimuli that could pose problems in interpreting these data as showing the cortical bases of object representation in the human visual cortex. For example, object stimuli are more structured than either sine wave gratings or texture images and may, therefore, capture attention more strongly than do grating stimuli. The object stimuli were, on average, more familiar to the subjects than the control stimuli (gratings or textures), which may make the object stimuli easier to process, allow them to capture attention, or make them more memorable.

More recent neuroimaging studies have attempted to create improved control stimuli for studying changes in rCBF associated with object processing. Schacter et al. (1995) presented subjects with images of possible and impossible objects (impossible objects were ones that could not exist in the 3-D world; they were structurally incoherent). As a baseline task, the subjects first saw the objects in a block of trials and had to respond with a keypress whenever an object disappeared from the screen. The subjects then studied 20 possible objects and 20 impossible objects; following this study phase, the subjects performed a possible/impossible judgment that required them to report whether the object was structurally coherent (i.e., had a real-world interpretation) or was structurally incoherent (i.e., did not have a real-world interpretation). The stimuli used in the possible/impossible judgment block were either old objects that had been observed in the study phase or new objects that had not been observed previously. Schacter et al. subtracted the blood flow changes in the baseline task from the blood flow changes in the possible/impossible task. Importantly, the same objects were presented in the baseline task, eliminating the problems associated with comparing different

stimuli (e.g., comparing gratings to objects). Schacter and colleagues reported increased blood flow to possible objects in the inferior temporal gyrus and fusiform gyrus. This activation was bilateral for old objects (i.e., objects observed in the study phase) and in the right hemisphere for new objects. These results seem compatible with the results of both Sergent et al. (1992) and Malach et al. (1995) in pointing to occipitotemporal areas, including the fusiform gyrus, as being involved in object representation. Schacter et al.'s results also indicate that there may be different rCBF changes associated with novel and familiar objects, with familiar objects being represented bilaterally in the extrastriate visual cortex. Other recent studies have also had subjects perform different tasks in order to localize blood flow changes associated with object recognition (see Köhler, Kapur, Moscovitch, Winocur, & Houle, 1995, who report changes in rCBF in the ventral posterior visual cortices bilaterally during an object identity-matching task).

Beyond manipulating the different tasks that subjects perform in neuroimaging studies (e.g., passive fixation vs. object recognition), Kanwisher and her colleagues have taken a different approach by manipulating the stimuli that subjects view (Kanwisher, Chun, McDermott, & Ledden, 1996; Kanwisher et al., 1997), an approach similar to that used by Malach et al. (1995). Kanwisher et al. (1997), for example, scanned subjects as they viewed pictures of familiar objects, pictures of novel objects, and scrambled stimuli, which were matched with the familiar objects on total luminance and number of pixels. Kanwisher and colleagues reasoned that these three stimulus types would receive different processing: (1) All three stimuli would engage feature extraction processes; (2) novel object and familiar object stimuli would also engage shape description processes that represented the objects' shape or structure; and (3) familiar objects would engage memory-matching processes. Subjects showed rCBF increases in the bilateral inferior occipitotemporal visual cortex when viewing both novel and familiar object stimuli but not when viewing scrambled stimuli. On the basis of these results, Kanwisher et al. (1997) concluded that this region of the occipitotemporal cortex is responsible for representing or describing an object's shape in a bottom-up manner (i.e., in a manner not related to the familiarity of the object). One puzzling aspect of these data, however, is that no *memory-matching* region was observed. That is, Kanwisher et al. (1997) did not find a region that responded differentially to familiar objects and novel objects. Although the failure to find a memory-matching region could have many causes (e.g., no explicit recognition task was used), one possibility is that shape description areas are also involved in memory for shapes. Thus, a single neural locus would be involved in storing familiar objects, and novel objects could be represented by partial matches to the stored, familiar objects. Such a proposal would be consistent with neural network models of object representation (see McClelland & Rumelhart, 1981, for a model that demonstrates partial matches for novel objects), and some neuroimaging evidence supports this: Novel objects often exhibit larger blood flow increases, as compared with familiar objects (see, e.g., Squire et al., 1992). This result can be explained by allowing novel objects to be partially matched to a number of stored familiar objects, allowing novel objects to partially activate a relatively large number of stored familiar object representations.

Although the majority of neuroimaging studies have focused on identifying those cortical regions associated with object representation, Kosslyn and colleagues have taken a different approach by studying different aspects of object processing. For example, as discussed previously in the Behavioral Results section , recognition may be influenced by the specific view of the object, according to viewer-centered theories of object representation. To study the role of orientation in object representation, Kosslyn et al. (1994) had subjects verify whether a spoken word matched a visually presented object while undergoing PET scans. The objects could appear either in a canonical orientation (e.g., a knife viewed from the side) or in a noncanonical orientation (e.g., a knife viewed from behind looking down the blade). Noncanonical objects resulted in larger blood flow changes in several regions, including bilaterally in the dorsolateral prefrontal cortex and bilaterally in the parietal cortex. The temporal lobe regions also demonstrated increased blood flow to the noncanonical views, as compared with the canonical views. Kosslyn et al. argue that the prefrontal regions are necessary to guide object recognition in a top-down manner when objects appear in noncanonical orientations. Kosslyn and colleagues conjecture that top-down guidance is required when viewing objects in noncanonical orientations because the bottom-up stimulus information is insufficient to permit flawless recognition; as a result, the bottom-up information can only be used to generate hypotheses about the object, and these hypotheses must be confirmed or refuted on the basis of higher level evaluation by frontal lobe mechanisms. Further specification of Kosslyn et al.'s framework will be required to determine whether the higher level evaluative processes involve transforming the image, as in viewer-centered theories (e.g., Tarr, 1995), or whether these evaluative processes involve extraction of parts that have been obscured by the noncanonical viewpoint, as might be suggested by some object-centered theories (e.g., Biederman, 1987).

Neuroimaging methods have added to our understanding of the role of the human extrastriate cortex in object representation. Although some might argue that neuroimaging studies only confirm what earlier neuropsychological studies revealed, some neuroimaging studies make novel contributions, such as the finding that different cortical areas can be activated with identical stimuli, depending on the task performed (see, e.g., Haxby et al., 1994). However, neuroimaging studies need to overcome two current shortcomings in order to integrate imaging results with the results from other methodologies. The first shortcoming is the role of human object

recognition areas (e.g., area LO and the fusiform gyrus) in computing the object invariances. Would human object recognition regions remain activated in the face of image-based changes (e.g., changes in size or retinal location)? The results from such studies could be difficult to interpret because of the stimulus differences required. However, such studies could specify further the computational processes that exist in human object recognition regions. The second shortcoming is the relationship between the regions activated in the human extrastriate cortex and the visual areas in the macaque temporal lobe. Understanding the homologies between areas in the human visual cortex and those in the macaque visual cortex could allow one to make predictions about the functional role of a region in the human extrastriate cortex on the basis of single-unit recordings or lesion studies.

Summary

Neuropsychology and neuroimaging have provided a wealth of information concerning human object recognition. These methods have indicated that multiple systems are probably involved in object representation and that other, nonvisual cortical areas may contribute to recognition. Despite the contributions of these methodologies, there are important reasons to constrain theories of object recognition with the methods discussed earlier in this review. For example, these methods do not provide information about the neural representations of objects; furthermore, the arch skeptic could argue that the anatomical localization offered by these methods is woefully inadequate when compared with anatomical studies with nonhuman primates. However, human neuropsychology and neuroimaging studies can provide an understanding of how object representations can be explicitly named or an understanding of the multiple object recognition systems that humans might possess.

The neuroimaging studies reviewed represent one of the most active areas in the cognitive neuroscience of object representation. As neuroimaging methods continue to develop, there will need to be additional emphasis on several problems that appear in this field. For example, the recent meta-analysis of object recognition imaging studies by Aguirre and Farah (1998) that appears in this volume suggests that not all neuroimaging studies converge. Instead, very different focal activations are produced, even when the same stimuli (e.g., words) are used in the experiments. Whether this lack of convergence is due to variability in brains or to variability in object recognition processes cannot be determined at present. Also, as Kanwisher et al. (1996) point out, many neuroimaging studies have confounds because the experimenters have manipulated both the stimuli used and the tasks performed by subjects (see Sergent et al., 1992, for an example). Clearly, either stimuli or tasks should be manipulated, or it may be even better to manipulate both task and stimuli in order to investigate interaction effects between stimuli and tasks. For example, novel objects and familiar objects may not produce different patterns of activation unless some type of recognition task is used. Both neuropsychology and neuroimaging methodologies have contributed to our understanding of primate object representation. The challenge for theorists using these methods will be to integrate their results with the computational and algorithmic levels of Marr's (1982) approach. Until such an integration occurs, progress in the field will be only piecemeal.

SUMMARY AND CONCLUSIONS

In this review, I have attempted to summarize some of the recent results from the object recognition literature. I have restricted my discussion to how different methodologies converge (or do not converge) on the role of the invariances in object representation. The invariances provide a nice thread that can be woven throughout the different disciplines that have investigated object representation. Specifically, there are excellent computational arguments for favoring object representation systems that compute size, spatial, and orientation invariances; such systems do not require an object representation to be stored for every possible size, spatial position, or orientation in which an object can appear. There is also strong evidence from psychophysical studies that indicates that the human recognition system computes these invariances (although the evidence is often mixed, as is especially the case for orientation invariance). Finally, neurophysiological, neuropsychological, and neuroimaging studies have also provided some evidence that suggests that the invariances can be computed by biological hardware, such as neurons in the inferior temporal lobe visual areas.

Although the invariances provide a coherent theme for a review of studies across disciplines, the invariances themselves do not provide a theory of the IT cortex, a theory of object recognition, or a computational mechanism by which object representation could occur. The invariances only provide a way of describing the outputs of the neural and computational mechanisms underlying object representation (see Plaut & Farah, 1990). The challenge across all disciplines involved in the study of object representation will be to explain how the invariances can arise from the IT cortex and what this neural region represents or computes. There have been several hypotheses concerning the role of the IT cortex in object representation, and these have been reviewed recently by Plaut and Farah.

One hypothesis about the neural mechanisms of object representation suggests that the relevant brain regions may be involved in storing simplified versions of the complex visual input or categorizing the visual input (see, e.g., Dean, 1982). For example, instead of storing the specific shape of a complex object such as a canary, the neural mechanisms of object representation may store a less precise or impoverished copy of this shape, but this impoverished representation would have many similarities with other stored instances of the category *bird*. This categorization process might result in object invariances be-

cause a large number of bird exemplars have been stored, allowing for the formation of a prototypical bird representation to emerge from the stored exemplars. This prototypical bird representation would contain the general properties of birds (e.g., bipedal, feathers, wings, etc.) that would allow this representation to be activated across a wide variety of retinal input (see Ratcliff & Newcombe, 1982). Another hypothesis that has been put forward is that the neural regions underlying object representation store a distributed-trace memory of objects (see, e.g., Gaffan, Harrison, & Gaffan, 1986a, 1986b). This distributed-trace memory involves representing a single object across a large number of neurons, each neuron coding for a different aspect of the object. For example, a canary may be represented in part by a neuron that codes for wing-shape attributes; this neuron could also be involved in representing other objects that have wings, such as cardinals or robins.

Hypotheses of the neural substrate of object representation, such as categorization or distributed-trace memory, are important because they tend to describe the processes underlying object representation, rather than merely describing what effects the object representation demonstrates (as with the invariances). However, the challenge to theorists focusing on object representation, as illustrated in this review, will be to integrate results across a large number of disciplines. Computational accounts will be important for specifying the function of the object representation system, but neurophysiology and other studies of the hardware level will place important constraints on computational accounts. Furthermore, many of the hypotheses put forth are unlikely to be mutually exclusive. For example, a categorization account may form prototypes via distributed-trace memories, an effect observed in artificial neural network models (see, e.g., McClelland & Rumelhart, 1985). Object representation mechanisms are likely to be performing many computations, so any theory that specifies a small number of operations, such as *categorization* or *prototype formation*, will probably fall short in providing an understanding of object processing in the brain. Like other reviewers, I have not endorsed a specific theory of object representation but have instead tried to point out some of the key results that will need to be accounted for by any theory of object representation.

REFERENCES

AGUIRRE, G., & FARAH, M. J. (1998). Human visual object recognition: What have we learned from neuroimaging? *Psychobiology*, **26**, 322-332.

ANDERSON, J. R. (1995). *Cognitive psychology and its implications* (4th ed.). New York: Freeman.

BARTRAM, D. J. (1974). The role of visual and semantic codes in object naming. *Cognitive Psychology*, **6**, 325-356.

BAY, E. (1953). Disturbances of visual perception and their examination. *Brain*, **76**, 515-550.

BENDER, M. B., & FELDMAN, M. (1972). The so-called "visual agnosias." *Brain*, **95**, 173-186.

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

BIEDERMAN, I., & COOPER, E. E. (1991). Evidence for complete transformational and reflectional invariance in human object priming. *Perception*, **20**, 585-593.

BIEDERMAN, I., & COOPER, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 121-133.

BIEDERMAN, I., & GERHARDSTEIN, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 1162-1182.

BIEDERMAN, I., & GERHARDSTEIN, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff (1995). *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1506-1514.

BIEDERMAN, I., & JU, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, **20**, 38-64.

BROOKS, R. A. (1981). Symbolic reasoning among 3-D models and 2-D images. *Artificial Intelligence*, **17**, 205-244.

BUNDESEN, C., & LARSEN, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 214-220.

DEAN, P. (1976). Effects of inferotemporal lesions on the behaviour of monkeys. *Psychological Bulletin*, **83**, 41-71.

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

DESIMONE, R., & GROSS, C. G. (1979). Visual areas in the temporal cortex of the macaque. *Brain Research*, **184**, 41-55.

DEYOE, E. A., CARMAN, G. J., BANDETTINI, P., GLICKMAN, S., WIESER, J., COX, R., MILLER, D., & NEITZ, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences*, **93**, 2382-2386.

DILL, M., & FAHLE, M. (1998). Limited translation invariance of human visual pattern recognition. *Perception & Psychophysics*, **60**, 65-81.

DUNCAN, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, **113**, 501-517.

EDELMAN, S. (1995). Representation, similarity, and the chorus of prototypes. *Minds & Machines*, **5**, 45-68.

EDELMAN, S. (1997). Computational theories of object recognition. *Trends in Cognitive Sciences*, **1**, 296-304.

EDELMAN, S., & BÜLTHOFF, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, **32**, 2385-2400.

EDELMAN, S., & DUVDEVANI-BAR, S. (1997). A model of visual recognition and categorization. *Philosophical Transactions of the Royal Society of London: Series B*, **352**, 1191-1202.

EDELMAN, S., & WEINSHALL, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics*, **64**, 209-219.

ELLIS, R., ALLPORT, D. A., HUMPHREYS, G. W., & COLLIS, J. (1989). Varieties of object constancy. *Quarterly Journal of Experimental Psychology*, **41A**, 775-796.

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

FARAH, M. J. (1991). Patterns of co-occurrence among associative agnosias: Implications for visual object representation. *Cognitive Neuropsychology*, **8**, 1-19.

FARAH, M. J., HAMMOND, K. M., LEVINE, D. N., & CALVANIO, R. (1988). Visual and spatial imagery: Dissociable systems of representation. *Cognitive Psychology*, **20**, 439-462.

FELLEMAN, D. J., & VAN ESSEN, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, **1**, 1-47.

FUJITA, I., TANAKA, K., ITO, M., & CHENG, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, **360**, 343-346.

GAFFAN, D., HARRISON, S., & GAFFAN, E. A. (1986a). Single and concurrent discrimination learning by monkeys after lesions of infero-

temporal cortex. *Quarterly Journal of Experimental Psychology*, **38B**, 31-51.

GAFFAN, D., HARRISON, S., & GAFFAN, E. A. (1986b). Visual identification following inferotemporal ablation in the monkey. *Quarterly Journal of Experimental Psychology*, **38B**, 5-30.

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

GROSS, C. G. (1992). Representation of visual stimuli in inferior temporal cortex. *Philosophical Transactions of the Royal Society of London: Series B*, **335**, 3-10.

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., & MISHKIN, M. (1977). The neural basis of stimulus equivalence across retinal translation. In S. Harnad, R. W. Doty, L. Goldstein, J. Jaynes, & G. Krauthamer (Eds.), *Lateralization in the nervous system* (pp. 109-122). 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.

HAXBY, J. V., HORWITZ, B., UNGERLEIDER, L. G., MAISOG, J. M., PIETRINI, P., & GRADY, C. L. (1994). The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations. *Journal of Neuroscience*, **14**, 6336-6353.

HINTON, G. E., MCCLELLAND, J. L., & RUMELHART, D. E. (1986). Distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 77-109). Cambridge, MA: MIT Press.

HOFFMAN, D. D., & RICHARDS, W. (1984). Parts of recognition. *Cognition*, **18**, 65-96.

HOLMES, E. J., & GROSS, C. G. (1984). Stimulus equivalence after inferior temporal lesions in monkeys. *Behavioral Neuroscience*, **98**, 898-901.

HUMMEL, J. E., & BIEDERMAN, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, **99**, 480-517.

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. (1984). Routes to object constancy: Implications from neurological impairments of object constancy. *Quarterly Journal of Experimental Psychology*, **36A**, 385-415.

HUMPHREYS, G. W., & RIDDOCH, M. J. (1987). The fractionation of visual agnosia. In G. W. Humphreys & M. J. Riddoch (Eds.), *Visual object processing: A cognitive neuropsychological approach* (pp. 281-306). Hove, U.K.: Erlbaum.

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.

JOLICOEUR, P. (1987). A size-congruency effect in memory for visual shape. *Memory & Cognition*, **15**, 531-543.

JOLICOEUR, P., GLUCK, M. A., & KOSSLYN, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, **16**, 243-275.

KANWISHER, N., CHUN, M. M., MCDERMOTT, J., & LEDDEN, P. J. (1996). Functional imaging of human visual recognition. *Cognitive Brain Research*, **5**, 55-67.

KANWISHER, N., WOODS, R. P., IACOBONI, M., & MAZZIOTTA, J. (1997). A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience*, **9**, 133-142.

KOBATAKE, E., & TANAKA, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, **71**, 856-867.

KÖHLER, S., KAPUR, S., MOSCOVITCH, M., WINOCUR, G., & HOULE, S. (1995). Dissociation of pathways for object and spatial vision: A PET study in humans. *NeuroReport*, **6**, 1865-1868.

KOSSLYN, S. M., ALPERT, N. M., THOMPSON, W. L., CHABRIS, C. F., RAUCH, S. L., & ANDERSON, A. K. (1994). Identifying objects seen from different viewpoints: A PET investigation. *Brain*, **117**, 1055-1071.

LARSEN, A., & BUNDESEN, C. (1978). Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 1-20.

LINDSAY, P. H., & NORMAN, D. A. (1977). *Human information processing* (2nd ed.). New York: Academic Press.

LISSAUER, H. (1988). A case of visual agnosia with a contribution to theory. *Cognitive Neuropsychology*, **5**, 157-192. (Originally published in 1890 as Ein Fall von Seelenblindheit nebst einem Beitrage zur Theorie derselben, *Archiv für Psychiatrie und Nervenkrankheiten*, **21**, 222-270)

LOGOTHETIS, N. K., & PAULS, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex*, **3**, 270-288.

LOGOTHETIS, N. K., PAULS, J., BÜLTHOFF, H. H., & POGGIO, T. (1994). View-dependent object recognition by monkeys. *Current Biology*, **4**, 401-414.

LOGOTHETIS, N. K., PAULS, J., & POGGIO, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, **5**, 552-563.

LOGOTHETIS, N. K., & SHEINBERG, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, **19**, 577-621.

LOWE, D. G. (1985). *Perceptual organization and visual recognition*. Boston: Kluwer.

LOWE, D. G. (1987a). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, **31**, 355-395.

LOWE, D. G. (1987b). The viewpoint consistency constraint. *International Journal of Computer Vision*, **1**, 57-72.

MALACH, R., REPPAS, J. B., BENSON, R. R., KWONG, K. K., JIANG, H., KENNEDY, W. A., LEDDEN, P. J., BRADY, T. J., ROSEN, B. R., & TOOTELL, R. B. H. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, **92**, 8135-8139.

MARR, D. (1982). *Vision*. San Francisco: Freeman.

MARR, D., & NISHIHARA, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B*, **204**, 301-328.

MCCARTHY, G., PUCE, A., GORE, J. C., & ALLISON, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, **9**, 605-610.

MCCLELLAND, J. L., & RUMELHART, D. E. (1981). An interactive model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, **88**, 375-407.

MCCLELLAND, J. L., & RUMELHART, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, **114**, 159-188.

MEL, B. (1996). SEEMORE: A view-based approach to 3-D object recognition using multiple visual cues. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing systems* (Vol. 8, pp. 865-871). Cambridge, MA: MIT Press.

MISHKIN, M. (1954). Visual discrimination performance following partial ablations of the temporal lobe: II. Ventral surface vs. hippocampus. *Journal of Comparative & Physiological Psychology*, **47**, 187-193.

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

MISHKIN, M., & PRIBRAM, K. H. (1954). Visual discrimination performance following partial ablations of the temporal lobe: I. Ventral vs. lateral. *Journal of Comparative & Physiological Psychology*, **47**, 14-20.

MIYASHITA, Y. (1993). Inferior temporal cortex: Where visual perception meets memory. *Annual Review of Neuroscience*, **16**, 245-263.

MOZER, M. C. (1991). *The perception of multiple objects: A connectionist approach*. Cambridge, MA: MIT Press.

NAKAMURA, H., GATTASS, R., DESIMONE, R., & UNGERLEIDER, L. G. (1993). The modular organization of projections from areas V1 and V2 to areas V4 and TEO in macaques. *Journal of Neuroscience*, **13**, 3681-3691.

NEISSER, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

PALMER, S. E., ROSCH, E., & CHASE, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 135-151). Hillsdale, NJ: Erlbaum.

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

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 of the Royal Society of London: Series B*, **223**, 293-317.

PETERSEN, S. E., FOX, P. T., POSNER, M. I., MINTUN, M., & RAICHLE, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, **331**, 585-589.

PETERSEN, S. E., FOX, P. T., SNYDER, A., & RAICHLE, M. E. (1990). Activation of prestriate and frontal cortical activity by words and word-like stimuli. *Science*, **249**, 1041-1044.

PINKER, S. (1984). Visual cognition: An introduction. *Cognition*, **18**, 1-63.

PLAUT, D. C., & FARAH, M. J. (1990). Visual object representation: Interpreting neurophysiological data within a computational framework. *Journal of Cognitive Neuroscience*, **2**, 320-343.

POGGIO, T., & EDELMAN, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, **343**, 263-266.

POSNER, M. I., & PETERSEN, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, **13**, 25-42.

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* (pp. 115-152). Pittsburgh: University of Pittsburgh.

PUCE, A., ALLISON, T., ASGARI, M., GORE, J. C., & MCCARTHY, G. (1996). Differential sensitivity of human visual cortex to faces, letter-strings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience*, **16**, 5205-5215.

RAO, S. C., RAINER, G., & MILLER, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science*, **276**, 821-824.

RATCLIFF, G., & NEWCOMBE, F. (1982). Object recognition: Some deductions from the clinical evidence. In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 147-171). New York: Academic Press.

ROCK, I., & DIVITA, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, **19**, 280-293.

ROCK, I., DIVITA, J., & BARBEITO, R. (1981). The effect on form perception of change of orientation in the third dimension. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 719-732.

ROCKLAND, K. S., & LUND, J. S. (1982). Widespread periodic intrinsic connections in the tree shrew visual cortex. *Science*, **215**, 1532-1534.

ROCKLAND, K. S., SALEEM, K. S., & TANAKA, K. (1994). Divergent feedback connections from areas V4 and TEO in the macaque. *Visual Neuroscience*, **11**, 579-600.

ROLAND, P. E., & FRIBERG, L. (1985). Localization of cortical areas activated by thinking. *Journal of Neurophysiology*, **53**, 1219-1243.

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.

RUBENS, A., & BENSON, D. F. (1971). Associative visual agnosia. *Archives of Neurology & Psychiatry*, **24**, 305-316.

SCHACTER, D. L., REIMAN, E., UECKER, A., POLSTER, M. R., YUN, L. S., & COOPER, L. A. (1995). Brain regions associated with retrieval of structurally coherent visual information. *Nature*, **376**, 587-590.

SCHWARTZ, E. L., DESIMONE, R., ALBRIGHT, T. D., & GROSS, C. G. (1983). Shape recognition and inferior temporal neurons. *Proceedings of the National Academy of Sciences*, **80**, 5776-5778.

SEACORD, L., GROSS, C. G., & MISHKIN, M. (1979). Role of inferior temporal cortex in interhemispheric transfer. *Brain Research*, **167**, 259-272.

SELFRIDGE, O. G., & NEISSER, U. (1960, August). Pattern recognition by machine. *Scientific American*, **203**, 60-68.

SERENO, M. I., DALE, A. M., REPPAS, J. B., KWONG, K. K., BELLIVEAU, J. W., BRADY, T. J., ROSEN, R. B., & TOOTELL, R. B. H. (1995). Borders of multiple visual areas revealed by functional magnetic resonance. *Science*, **268**, 889-893.

SERGENT, J., OHTA, S., & MACDONALD, B. (1992). Functional neuroanatomy of face and object processing. *Brain*, **115**, 15-36.

SHEPARD, R. N., & COOPER, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.

SHEPARD, R. N., & METZLER, J. (1971). Mental rotation of three-dimensional objects. *Science*, **171**, 701-703.

SQUIRE, L. R., OJEMANN, J. G., MIEZIN, F. M., PETERSEN, S. E., VIDEEN, T. O., & RAICHLE, M. E. (1992). Activation of the hippocampus in normal humans: A functional anatomical study of memory. *Proceedings of the National Academy of Sciences*, **89**, 1837-1841.

TANAKA, K. (1992). Inferotemporal cortex and higher visual functions. *Current Opinion in Neurobiology*, **2**, 502-505.

TANAKA, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, **19**, 109-139.

TANAKA, K. (1997). Mechanisms of visual object recognition: Monkey and human studies. *Current Opinion in Neurobiology*, **7**, 523-527.

TANAKA, K. SAITO, H., FUKADA, Y., & MORIYA, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, **66**, 170-189.

TARR, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review*, **2**, 55-82.

TARR, M. J., & BÜLTHOFF, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1494-1505.

TARR, M. J., & PINKER, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, **21**, 233-282.

TARR, M. J., & PINKER, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science*, **1**, 253-256.

TEUBER, H. L. (1968). Perception. In L. Weiskrantz (Ed.), *Analysis of behavioral change* (pp. 274-328). New York: Harper & Row.

TURNBULL, O. H. (1997). A double dissociation between knowledge of object identity and object orientation. *Neuropsychologia*, **35**, 567-570.

ULLMAN, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, **32**, 193-254.

ULLMAN, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA: MIT Press.

UNGERLEIDER, L. G., GANZ, L., & PRIBRAM, K. H. (1977). Size constancy in rhesus monkeys: Effects of pulvinar, prestriate, and inferotemporal lesions. *Experimental Brain Research*, **27**, 251-269.

UNGERLEIDER, L. G., & MISHKIN, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press.

VAN ESSEN, D. C., & DRURY, H. A. (1997). Structural and functional analyses of human cerebral cortex using a surface-based atlas. *Journal of Neuroscience*, **17**, 7079-7102.

VECERA, S. P., & FARAH, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, **123**, 146-160.

VON BONIN, G., & BAILEY, P. (1950). *The neocortex of the chimpanzee*. Urbana: University of Illinois Press.

WANG, G., TANAKA, K., & TANIFUJI, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science*, **272**, 1665-1668.

WAPNER, W., JUDD, T., & GARDNER, H. (1978). Visual agnosia in an artist. *Cortex*, **14**, 343-364.

WARRINGTON, E. K. (1985). Agnosia: The impairment of object recognition. In P. J. Vinken, G. W. Bruyn, & H. L. Klawans (Eds.), *Handbook of clinical neurology: Vol. 1. Clinical neuropsychology* (pp. 333-349). Amsterdam: Elsevier.

WEISKRANTZ, L., & SAUNDERS, R. C. (1984). Impairments of visual object transforms in monkeys. *Brain*, **107**, 1033-1072.

YIN, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, **81**, 141-145.

YOUNG, M. P. (1995). Open questions about the neural mechanisms of visual pattern recognition. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 463-474). Cambridge, MA: MIT Press.

NOTE

1. For the remainder of the paper I will use *primate visual system* to refer to both humans and nonhuman primates. When referring to either humans or nonhuman primates, I will specify which group I intend to discuss.

(Manuscript received August 4, 1998;
accepted for publication August 19, 1998.)