Psychological Review
1987, Vol. 94, No. 2, 115–147

0033-295X/87/$00.75

Recognition-by-Components: A Theory of Human Image Understanding

Irving Biederman
State University of New York at Buffalo

The perceptual recognition of objects is conceptualized to be a process in which the image of the input is segmented at regions of deep concavity into an arrangement of simple geometric components, such as blocks, cylinders, wedges, and cones. The fundamental assumption of the proposed theory, recognition-by-components (RBC), is that a modest set of generalized-cone components, called geons ($N < 36$), can be derived from contrasts of five readily detectable properties of edges in a two-dimensional image: curvature, collinearity, symmetry, parallelism, and cotermination. The detection of these properties is generally invariant over viewing position and image quality and consequently allows robust object perception when the image is projected from a novel viewpoint or is degraded. RBC thus provides a principled account of the heretofore undecided relation between the classic principles of perceptual organization and pattern recognition: The constraints toward regularization (Pragnanz) characterize not the complete object but the object's components. Representational power derives from an allowance of free combinations of the geons. A Principle of Componential Recovery can account for the major phenomena of object recognition: If an arrangement of two or three geons can be recovered from the input, objects can be quickly recognized even when they are occluded, novel, rotated in depth, or extensively degraded. The results from experiments on the perception of briefly presented pictures by human observers provide empirical support for the theory.

Any single object can project an infinity of image configurations to the retina. The orientation of the object to the viewer can vary continuously, each giving rise to a different two-dimensional projection. The object can be occluded by other objects or texture fields, as when viewed behind foliage. The object need not be presented as a full-colored textured image but instead can be a simplified line drawing. Moreover, the object can even be missing some of its parts or be a novel exemplar of its particular category. But it is only with rare exceptions that an image fails to be rapidly and readily classified, either as an instance of a familiar object category or as an instance that cannot be so classified (itself a form of classification).

A Do-It-Yourself Example

Consider the object shown in Figure 1. We readily recognize it as one of those objects that cannot be classified into a familiar category. Despite its overall unfamiliarity, there is near unanimity in its descriptions. We parse—or segment—its parts at regions of deep concavity and describe those parts with common, simple volumetric terms, such as "a block," "a cylinder," "a funnel or truncated cone." We can look at the zig-zag horizontal brace as a texture region or zoom in and interpret it as a series of connected blocks. The same is true of the mass at the lower left: we can see it as a texture area or zoom in and parse it into its various bumps.

Although we know that it is not a familiar object, after a while we can say what it resembles: "A New York City hot dog cart, with the large block being the central food storage and cooking area, the rounded part underneath as a wheel, the large arc on the right as a handle, the funnel as an orange juice squeezer and the various vertical pipes as vents or umbrella supports." It is not a good cart, but we can see how it might be related to one. It is like a 10-letter word with 4 wrong letters.

We readily conduct the same process for any object, familiar or unfamiliar, in our foveal field of view. The manner of segmentation and analysis into components does not appear to depend on our familiarity with the particular object being identified.

The naive realism that emerges in descriptions of nonsense objects may be reflecting the workings of a representational system by which objects are identified.

An Analogy Between Speech and Object Perception

As will be argued in a later section, the number of categories into which we can classify objects rivals the number of words that can be readily identified when listening to speech. Lexical access during speech perception can be successfully modeled as a process mediated by the identification of individual primitive elements, the phonemes, from a relatively small set of primitives (Marslen-Wilson, 1980). We only need about 44 phonemes to code all the words in English, 15 in Hawaiian, 55 to represent virtually all the words in all the languages spoken around the world. Because the set of primitives is so small and each pho-

This research was supported by the Air Force Office of Scientific Research Grants F4962083C0086 and 86-0106.

I would like to express my deep appreciation to Thomas W. Blickle and Ginny Ju for their invaluable contributions to all phases of the empirical research described in this article. Thanks are also due to Elizabeth A. Beiring, John Clapper, and Mary Lloyd for their assistance in the conduct of the experimental research. Discussions I had with James R. Pomerantz, John Artim, and Brian Fisher helped improve aspects of the manuscript.

Correspondence concerning this article should be addressed to Irving Biederman, Department of Psychology, State University of New York at Buffalo, Park Hall, Amherst, New York 14260.

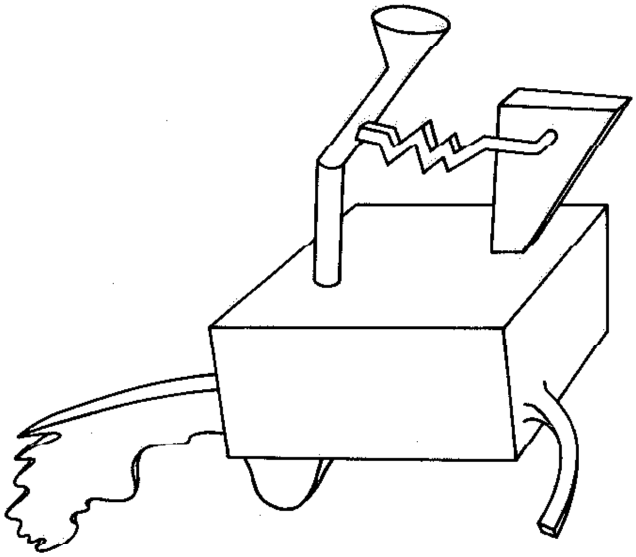


Figure 1. A do-it-yourself object. (There is strong consensus in the segmentation loci of this configuration and in the description of its parts.)

neme specifiable by dichotomous (or trichotomous) contrasts (e.g., voiced vs. unvoiced, nasal vs. oral) on a handful of attributes, one need not make particularly fine discriminations in the speech stream. The representational power of the system derives from its permissiveness in allowing relatively free combinations of its primitives.

The hypothesis explored here is that a roughly analogous system may account for our capacities for object recognition. In the visual domain, however, the primitive elements would not be phonemes but a modest number of simple geometric components—generally convex and volumetric—such as cylinders, blocks, wedges, and cones. Objects are segmented, typically at regions of sharp concavity, and the resultant parts matched against the best fitting primitive. The set of primitives derives from combinations of contrasting characteristics of the edges in a two-dimensional image (e.g., straight vs. curved, symmetrical vs. asymmetrical) that define differences among a set of simple volumes (viz., those that tend to be symmetrical and lack sharp concavities). As in speech perception, these contrasts need only be dichotomous or trichotomous rather than quantitative, so that the human's limited capacities for absolute judgment are not taxed. The particular properties of edges that are postulated to be relevant to the generation of the volumetric primitives have the desirable properties that they are invariant over changes in orientation and can be determined from just a few points on each edge. Consequently, they allow a primitive to be extracted with great tolerance for variations of viewpoint, occlusion, and noise.

Just as the relations among the phonemes are critical in lexical access—"fur" and "rough" have the same phonemes but are not the same words—the relations among the volumes are critical for object recognition: Two different arrangements of the same components could produce different objects. In both cases, the representational power derives from the enormous number of combinations that can arise from a modest number of primitives. The relations in speech are limited to left-to-right (sequential) orderings; in the visual domain a richer set of possible relations allows a far greater representational capacity from a comparable number of primitives. The matching of objects in recognition is hypothesized to be a process in which the perceptual input is matched against a representation that can be described by a few simple categorized volumes in specified relations to each other.

Theoretical Domain: Primal Access to Contour-Based Perceptual Categories

Our theoretical goal is to account for the initial categorization of isolated objects. Often, but not always, this categorization will be at a basic level, for example, when we know that a given object is a typewriter, a banana, or a giraffe (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Much of our knowledge about objects is organized at this level of categorization: the level at which there is typically some readily available name to describe that category (Rosch et al., 1976). The hypothesis explored here predicts that when the componential description of a particular subordinate differs substantially from a basic-level prototype, that is, when a subordinate is perceptually nonprototypical, categorizations will initially be made at the subordinate level. Thus, we might know that a given object is a floor lamp, a penguin, a sports car, or a dachshund more rapidly than we know that it is a lamp, a bird, a car, or a dog (e.g., Jolicoeur, Gluck, & Kosslyn, 1984). (For both theoretical and expository purposes, these readily identifiable nonprototypical members [subordinates] of basic level categories will also be considered basic level in this article.)

Count Versus Mass Noun Entities: The Role of Surface Characteristics

There is a restriction on the scope of this approach of volumetric modeling that should be noted. The modeling has been limited to concrete entities with specified boundaries. In English, such objects are typically designated by count nouns. These are concrete objects that have specified boundaries and to which we can apply the indefinite article and number. For example, for a count noun such as "chair" we can say "a chair" or "three chairs." By contrast, mass nouns are concrete entities to which the indefinite article or number cannot be applied, such as water, sand, or snow. So we cannot say "a water" or "three sands," unless we refer to a count noun shape, as in "a drop of water," "a bucket of water," "a grain of sand," or "a snowball," each of which does have a simple volumetric description. We conjecture that mass nouns are identified primarily through surface characteristics such as texture and color, rather than through volumetric primitives.

Primal Access

Under restricted viewing and uncertain conditions, as when an object is partially occluded, texture, color, and other cues (such as position in the scene and labels) may constitute part or all of the information determining memory access, as for example when we identify a particular shirt in the laundry pile from seeing just a bit of fabric. Such identifications are indirect, typically the result of inference over a limited set of possible objects. (Additional analyses of the role of surface features is presented later in the discussion of the experimental comparison of the perceptibility of color photography and line drawings.) The goal of the present effort is to account for what can be called *primal access:* the first contact of a perceptual input from an isolated, unanticipated object to a representation in memory.

Basic Phenomena of Object Recognition

Independent of laboratory research, the phenomena of everyday object identification provide strong constraints on possible models of recognition. In addition to the fundamental phenomenon that objects can be recognized at all (not an altogether obvious conclusion), at least five facts are evident. Typically, an object can be recognized rapidly, when viewed most from novel orientations, under moderate levels of visual noise, when partially occluded, and when it is a new exemplar of a category.

The preceding five phenomena constrain theorizing about object interpretation in the following ways:

1. Access to the mental representation of an object should not be dependent on absolute judgments of quantitative detail, because such judgments are slow and error prone (Garner, 1962; Miller, 1956). For example, distinguishing among just several levels of the degree of curvature or length of an object typically requires more time than that required for the identification of the object itself. Consequently, such quantitative processing cannot be the controlling factor by which recognition is achieved.
2. The information that is the basis of recognition should be relatively invariant with respect to orientation and modest degradation.
3. Partial matches should be computable. A theory of object interpretation should have some principled means for computing a match for occluded, partial, or new exemplars of a given category. We should be able to account for the human's ability to identify, for example, a chair when it is partially occluded by other furniture, or when it is missing a leg, or when it is a new model.

Recognition-by-Components: An Overview

Our hypothesis, recognition-by-components (RBC), bears some relation to several prior conjectures for representing objects by parts or modules (e.g., Binford, 1971; Brooks, 1981; Guzman, 1971; Marr, 1977; Marr & Nishihara, 1978; Tversky & Hemenway, 1984). RBC's contribution lies in its proposal for a particular vocabulary of components derived from perceptual mechanisms and its account of how an arrangement of these components can access a representation of an object in memory.

Stages of Processing

Figure 2 presents a schematic of the presumed subprocesses by which an object is recognized. These stages are assumed to be arranged in cascade. An early edge extraction stage, responsive to differences in surface characteristics namely, luminance, texture, or color, provides a line drawing description of the object. From this description, nonaccidental properties of image edges (e.g., collinearity, symmetry) are detected. Parsing is performed, primarily at concave regions, simultaneously with a detection of nonaccidental properties. The nonaccidental properties of the parsed regions provide critical constraints on the identity of the components. Within the temporal and contextual constraints of primal access, the stages up to and including the identification of components are assumed to be bottom-up.[1] A delay in the determination of an object's components should have a direct effect on the identification latency of the object.

The arrangement of the components is then matched against a representation in memory. It is assumed that the matching of the components occurs in parallel, with unlimited capacity. Partial matches are possible with the degree of match assumed to be proportional to the similarity in the components between the image and the representation.[2] This stage model is presented to provide an overall theoretical context. The focus of this article is on the nature of the units of the representation.

When an image of an object is painted on the retina, RBC assumes that a representation of the image is segmented—or parsed—into separate regions at points of deep concavity, particularly at cusps where there are discontinuities in curvature (Marr & Nishihara, 1978). In general, paired concavities will arise whenever convex volumes are joined, a principle that Hoffman and Richards (1985) term *transversality.* Such segmentation conforms well with human intuitions about the boundaries of object parts and does not depend on familiarity

[1] The only top-down route shown in Figure 2 is an effect of the nonaccidental properties on edge extraction. Even this route (aside from collinearity and smooth curvature) would run counter to the desires of many in computational vision (e.g., Marr, 1982) to build a completely bottom-up system for edge extraction. This assumption was developed in the belief that edge extraction does not depend on prior familiarity with the object. However, as with the nonaccidental properties, a top-down route from the component determination stage to edge extraction could precede independent of familiarity with the object itself. It is possible that an edge extraction system with a competence equivalent to that of a human—an as yet unrealized accomplishment—will require the inclusion of such top-down influences. It is also likely that other top-down routes, such as those from expectancy, object familiarity, or scene constraints (e.g., Biederman, 1981; Biederman, Mezzanotte, & Rabinowitz, 1982), will be observed at a number of the stages, for example, at segmentation, component definition, or matching, especially if edges are degraded. These have been omitted from Figure 2 in the interests of simplicity and because their actual paths of influence are as yet undetermined. By proposing a general account of object recognition, it is hoped that the proposed theory will provide a framework for a principled analysis of top-down effects in this domain.

[2] Modeling the matching of an object image to a mental representation is a rich, relatively neglected problem area. Tversky's (1977) contrast model provides a useful framework with which to consider this similarity problem in that it readily allows distinctive features (components) of the image to be considered separately from the distinctive components of the representation. This allows principled assessments of similarity for partial objects (components in the representation but not in the image) and novel objects (containing components in the image that are not in the representation). It may be possible to construct a dynamic model based on a parallel distributed process as a modification of the kind proposed by McClelland and Rumelhart (1981) for word perception, with components playing the role of letters. One difficulty of such an effort is that the set of neighbors for a given word is well specified and readily available from a dictionary; the set of neighbors for a given object is not.

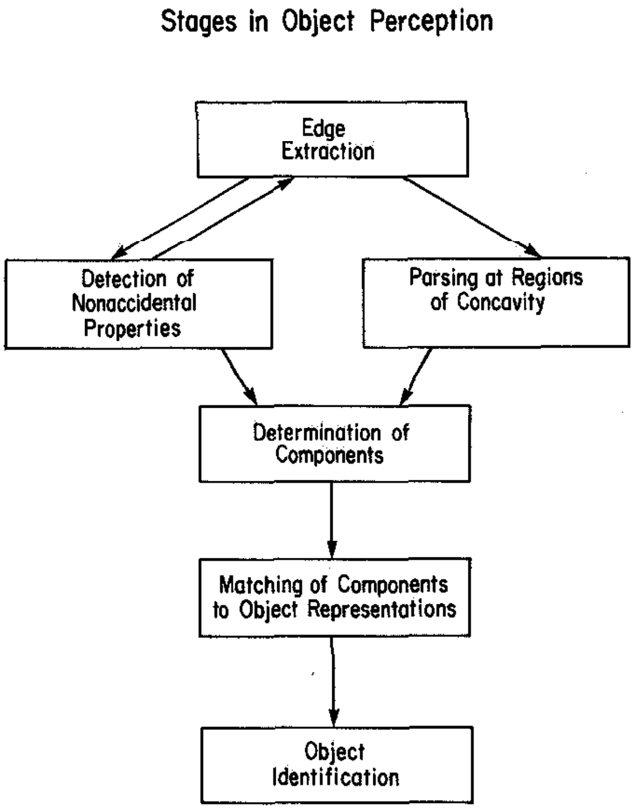

Stages in Object Perception
Edge Extraction
Detection of Nonaccidental Properties
Parsing at Regions of Concavity
Determination of Components
Matching of Components to Object Representations
Object Identification

Figure 2. Presumed processing stages in object recognition.

with the object, as was demonstrated with the nonsense object in Figure 1.

Each segmented region is then approximated by one of a possible set of simple components, called *geons* (for "geometrical ions"), that can be modeled by generalized cones (Binford, 1971; Marr, 1977, 1982). A generalized cone is the volume swept out by a cross section moving along an axis (as illustrated in Figure 5). (Marr [1977, 1982] showed that the contours generated by any smooth surface could be modeled by a generalized cone with a convex cross section.) The cross section is typically hypothesized to be at right angles to the axis. Secondary segmentation criteria (and criteria for determining the axis of a component) are those that afford descriptions of volumes that maximize symmetry, axis length, and constancy of the size and curvature of the cross section of the component. Of these, symmetry often provides the most compelling subjective basis for selecting subparts (Brady & Asada, 1984; Connell, 1985). These secondary bases for segmentation and component identification are discussed below.

The primitive components are hypothesized to be simple, typically symmetrical volumes lacking sharp concavities, such as blocks, cylinders, spheres, and wedges. The fundamental perceptual assumption of RBC is that the components can be differentiated on the basis of perceptual properties in the two-dimensional image that are readily detectable and relatively independent of viewing position and degradation. These perceptual properties include several that traditionally have been thought of as principles of perceptual organization, such as good continuation, symmetry, and Pragnanz. RBC thus provides a principled account of the relation between the classic phenomena of perceptual organization and pattern recognition: Although objects can be highly complex and irregular, the units by which objects are identified are simple and regular. The constraints toward regularization (Pragnanz) are thus assumed to characterize not the complete object but the object's components.

Color and Texture

The preceding account is clearly edge-based. Surface characteristics such as color, brightness, and texture will typically have only secondary roles in primal access. This should not be interpreted as suggesting that the perception of surface characteristics per se is delayed relative to the perception of the components (but see Barrow & Tenenbaum, 1981), but merely that in most cases the surface characteristics are generally less efficient routes for accessing the classification of a count object. That is, we may know that a chair has a particular color and texture simultaneously with its componential description, but it is only the volumetric description that provides efficient access to the mental representation of "chair."[3]

Relations Among the Components

Although the components themselves are the focus of this article, as noted previously the arrangement of primitives is necessary for representing a particular object. Thus, an arc side-connected to a cylinder can yield a cup, as shown in Figure 3C. Different arrangements of the same components can readily lead to different objects, as when an arc is connected to the top of the cylinder to produce a pail (Figure 3D). Whether a component is attached to a long or short surface can also affect classification, as with the arc producing either an attaché case (Figure 3A) or a strongbox (Figure 3B).

The identical situation between primitives and their arrangement exists in the phonemic representation of words, where a given subset of phonemes can be rearranged to produce different words.

The representation of an object would thus be a structural description that expressed the relations among the components (Ballard & Brown, 1982; Winston, 1975). A suggested (minimal) set of relations will be described later (see Table 1). These

[3] There are, however, objects that would seem to require both a volumetric description and a texture region for an adequate representation, such as hairbrushes, typewriter keyboards, and corkscrews. It is unlikely that many of the individual bristles, keys, or coils are parsed and identified prior to the identification of the object. Instead those regions are represented through the statistical processing that characterizes their texture (for example, Beck, Prazdny, & Rosenfeld, 1983; Julesz, 1981), although we retain a capacity to zoom down and attend to the volumetric nature of the individual elements. The structural description that would serve as a representation of such objects would include a statistical specification of the texture field along with a specification of the larger volumetric components. These compound texture–componential objects have not been studied, but it is possible that the characteristics of their identification would differ from objects that are readily defined solely by their arrangement of volumetric components.

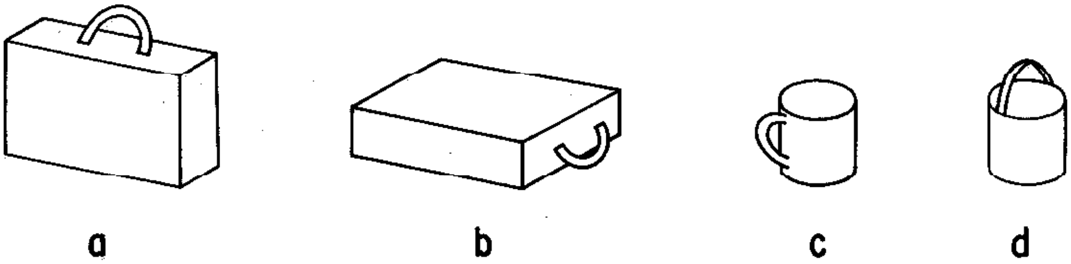


Figure 3. Different arrangements of the same components can produce different objects.

relations include specification of the relative sizes of the components, their orientation and the locus of their attachment.

Nonaccidental Properties: A Perceptual Basis for a Componential Representation

Recent theoretical analyses of perceptual organization (Binford, 1981; Lowe, 1984; Rock, 1983; Witkin & Tenenbaum, 1983) provide a perceptual basis for generating a set of geons. The central organizational principle is that certain properties of edges in a two-dimensional image are taken by the visual system as strong evidence that the edges in the three-dimensional world contain those same properties. For example, if there is a straight line in the image (*collinearity*), the visual system infers that the edge producing that line in the three-dimensional world is also straight. The visual system ignores the possibility that the property in the image might be a result of a (highly unlikely) accidental alignment of eye and curved edge. Smoothly curved elements in the image (*curvilinearity*) are similarly inferred to arise from smoothly curved features in the three-dimensional world. These properties, and the others described later, have been termed *nonaccidental* (Witkin & Tenenbaum, 1983) in that they would only rarely be produced by accidental alignments of viewpoint and object features and consequently are generally unaffected by slight variations in viewpoint.

If the image is symmetrical (*symmetry*), we assume that the object projecting that image is also symmetrical. The order of symmetry is also preserved: Images that are symmetrical under both reflection and 90° increments of rotation, such as a square or circle, are interpreted as arising from objects (or surfaces) that are symmetrical under both rotation and reflection. Although skew symmetry is often readily perceived as arising from a tilted symmetrical object or surface (Palmer, 1983), there are cases where skew symmetry is not readily detected (Attneave, 1982). When edges in the image are parallel or coterminate we assume that the real-world edges also are parallel or coterminate, respectively.

These five nonaccidental properties and the associated three-dimensional inferences are described in Figure 4 (adapted from Lowe, 1984). Witkin and Tenenbaum (1983; see also Lowe, 1984) argue that the leverage provided by the nonaccidental relations for inferring a three-dimensional structure from a two-dimensional image edges is so great as to pose a challenge to the effort in computational vision and perceptual psychology that assigned central importance to variation in local surface characteristics, such as luminance gradients, from which surface curvature could be determined (as in Besl & Jain, 1986). Although a surface property derived from such gradients will be invariant over some transformations, Witkin and Tenenbaum (1983) demonstrate that the suggestion of a volumetric component through the shape of the surface's silhouette can readily override the perceptual interpretation of the luminance gradient. The psychological literature, summarized in the next section, provides considerable evidence supporting the assumption that these nonaccidental properties can serve as primary organizational constraints in human image interpretation.

Psychological Evidence for the Rapid Use of Nonaccidental Relations

There can be little doubt that images are interpreted in a manner consistent with the nonaccidental principles. But are these relations used quickly enough to provide a perceptual basis for the components that allow primal access? Although all the principles have not received experimental verification, the available evidence strongly suggests an affirmative answer to the preceding question. There is strong evidence that the visual system quickly assumes and uses collinearity, curvature, symmetry, and cotermination. This evidence is of two sorts: (a) demonstrations, often compelling, showing that when a given two-dimensional relation is produced by an accidental alignment of object and image, the visual system accepts the relation as existing in the three-dimensional world; and (b) search tasks showing that when a target differs from distractors in a nonaccidental property, as when one is searching for a curved arc among straight segments, the detection of that target is facilitated compared to conditions where targets and background do not differ in such properties.

Collinearity versus curvature. The demonstration of the collinearity or curvature relations is too obvious to be performed as an experiment. When looking at a straight segment, no observer would assume that it is an accidental image of a curve. That the contrast between straight and curved edges is readily available for perception was shown by Neisser (1963). He found that a search for a letter composed only of straight segments, such as a Z, could be performed faster when in a field of curved distractors, such as C, G, O, and Q, then when among other letters composed of straight segments such as N, W, V, and M.

Symmetry and parallelism. Many of the Ames demonstrations (Ittleson, 1952), such as the trapezoidal window and Ames room, derive from an assumption of symmetry that includes parallelism. Palmer (1980) showed that the subjective directionality of arrangements of equilateral triangles was based on the

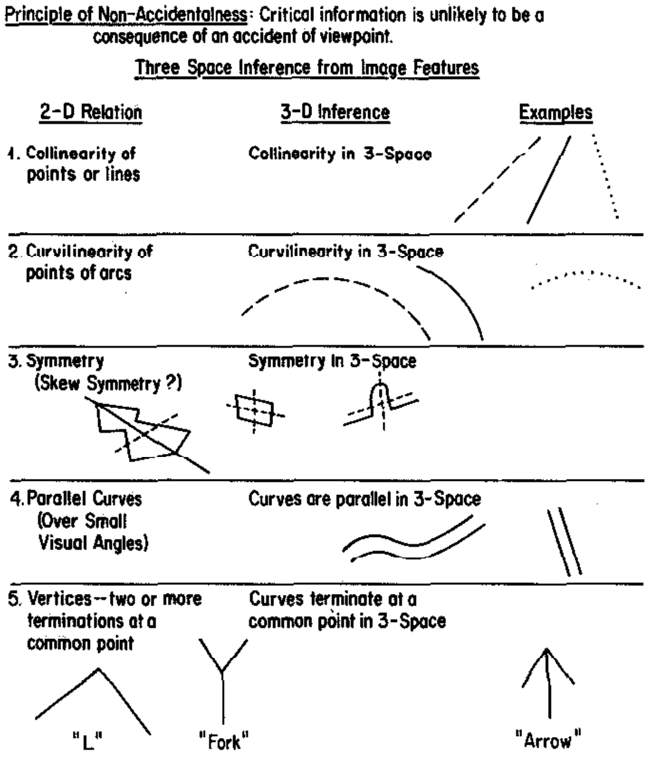

Principle of Non-Accidentalness: Critical information is unlikely to be a consequence of an accident of viewpoint.
Three Space Inference from Image Features
2-D Relation
3-D Inference
Examples
1. Collinearity of points or lines
Collinearity in 3-Space
2. Curvilinearity of points of arcs
Curvilinearity in 3-Space
3. Symmetry (Skew Symmetry ?)
Symmetry in 3-Space
4. Parallel Curves (Over Small Visual Angles)
Curves are parallel in 3-Space
5. Vertices--two or more terminations at a common point
Curves terminate at a common point in 3-Space
"L"
"Fork"
"Arrow"

Figure 4. Five nonaccidental relations. (From Figure 5.2, *Perceptual organization and visual recognition* [p. 77] by David Lowe. Unpublished doctorial dissertation, Stanford University. Adapted by permission.)

derivation of an axis of symmetry for the arrangement. King, Meyer, Tangney, and Biederman (1976) demonstrated that a perceptual bias toward symmetry contributed to apparent shape constancy effects. Garner (1974), Checkosky and Whitlock (1973), and Pomerantz (1978) provided ample evidence that not only can symmetrical shapes be quickly discriminated from asymmetrical stimuli, but that the degree of symmetry was also a readily available perceptual distinction. Thus, stimuli that were invariant under both reflection and 90° increments in rotation could be rapidly discriminated from those that were only invariant under reflection (Checkosky & Whitlock, 1973).

Cotermination. The "peephole perception" demonstrations, such as the Ames chair (Ittleson, 1952) or the physical realization of the "impossible" triangle (Penrose & Penrose, 1958), are produced by accidental alignment of the ends of noncoterminous segments to produce—from one viewpoint only—L, Y, and arrow vertices. More recently, Kanade (1981) has presented a detailed analysis of an "accidental" chair of his own construction. The success of these demonstrations document the immediate and compelling impact of cotermination.

The registration of cotermination is important for determining vertices, which provide information that can serve to distinguish the components. In fact, one theorist (Binford, 1981) has suggested that the major function of eye movements is to determine coincidence of segments. "Coincidence" would include not only cotermination of edges but the termination of one edge on another, as with a T vertex. With polyhedra (volumes produced by planar surfaces), the Y, arrow, and L vertices allow inference as to the identity of the volume in the image. For example, the silhouette of a brick contains a series of six vertices, which alternate between Ls and arrows, and an internal Y vertex, as illustrated in Figure 5. The Y vertex is produced by the cotermination of three segments, with none of the angles greater than 180°. (An arrow vertex, also formed from the cotermination of three segments, contains an angle that exceeds 180°; an L vertex is formed by the cotermination of two segments.) As shown in Figure 5, this vertex is not present in components that have curved cross sections, such as cylinders, and thus can provide a distinctive cue for the cross-section edge. (The curved Y vertex present in a cylinder can be distinguished from the Y or arrow vertices in that the termination of one segment in the curved Y is tangent to the other segment [Chakravarty, 1979].)

Perkins (1983) has described a perceptual bias toward parallelism in the interpretation of this vertex.[4] Whether the presence of this particular internal vertex can facilitate the identification of a brick versus a cylinder is not yet known, but a recent study by Biederman and Blickle (1985), described below, demonstrated that deletion of vertices adversely affected object recognition more than deletion of the same amount of contour at midsegment.

The T vertex represents a special case in that it is not a locus of cotermination (of two or more segments) but only the termination of one segment on another. Such vertices are important for determining occlusion and thus segmentation (along with concavities), in that the edge forming the (normally) vertical segment of the T cannot be closer to the viewer than the segment forming the top of the T (Binford, 1981). By this account, the T vertex might have a somewhat different status than the Y, arrow, and L vertices, in that the T's primary role would be in segmentation, rather than in establishing the identity of the volume.[5]

Vertices composed of three segments, such as the Y and ar-

[4] When such vertices formed the central angle in a polyhedron, Perkins (1983) reported that the surfaces would almost always be interpreted as meeting at right angles, as long as none of the three angles was less than 90°. Indeed, such vertices cannot be projections of acute angles (Kanade, 1981) but the human appears insensitive to the possibility that the vertices could have arisen from obtuse angles. If one of the angles in the central Y vertex was acute, then the polyhedra would be interpreted as irregular. Perkins found that subjects from rural areas of Botswana, where there was a lower incidence of exposure to carpentered (right-angled) environments, had an even stronger bias toward rectilinear interpretations than did Westerners (Perkins & Deregowski, 1982).

[5] The arrangement of vertices, particularly for polyhedra, offers constraints on "possible" interpretations of lines as convex, concave, or occluding (e.g., Sugihara, 1984). In general, the constraints take the form that a segment cannot change its interpretation, for example, from concave to convex, unless it passes through a vertex. "Impossible" objects can be constructed from violations of this constraint (Waltz, 1975) as well as from more general considerations (Sugihara, 1982, 1984). It is tempting to consider that the visual system captures these constraints in the way in which edges are grouped into objects, but the evidence would seem to argue against such an interpretation. The impossibility of most impossible objects is not immediately registered, but requires scrutiny and thought before the inconsistency is detected. What this means in the present context is that the visual system has a capacity for classifying vertices locally, but no perceptual routines for determining the global consistency of a set of vertices.

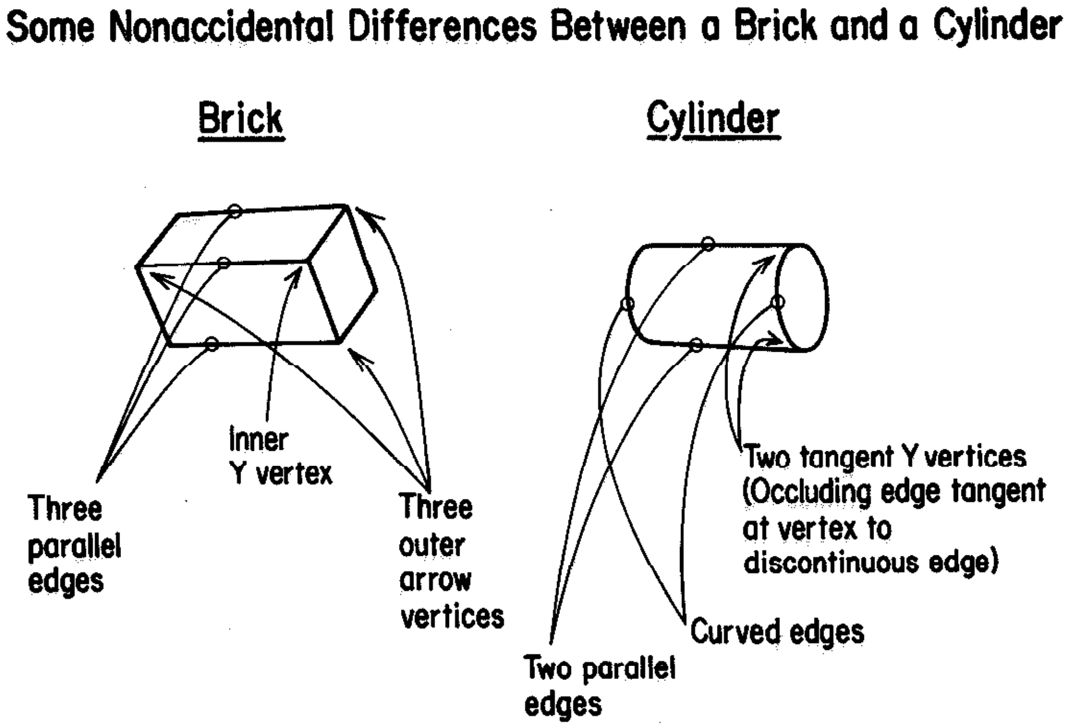

Some Nonaccidental Differences Between a Brick and a Cylinder
Brick
Cylinder
Three parallel edges
Inner Y vertex
Three outer arrow vertices
Two parallel edges
Curved edges
Two tangent Y vertices (Occluding edge tangent at vertex to discontinuous edge)

Figure 5. Some differences in nonaccidental properties between a cylinder and a brick.

row, and their curved counterparts, are important determinants as to whether a given component is volumetric or planar. Planar components (to be discussed later) lack three-pronged vertices.

The high speed and accuracy of determining a given nonaccidental relation (e.g., whether some pattern is symmetrical) should be contrasted with performance in making absolute quantitative judgments of variations in a single physical attribute, such as length of a segment or degree of tilt or curvature. For example, the judgment as to whether the length of a given segment is 10, 12, 14, 16, or 18 cm is notoriously slow and error prone (Beck, Prazdny, & Rosenfeld, 1983; Fildes & Triggs, 1985; Garner, 1962; Miller, 1956; Virsu, 1971a, 1971b). Even these modest performance levels are challenged when the judgments have to be executed over the brief 100-ms intervals (Egeth & Pachella, 1969) that are sufficient for accurate object identification. Perhaps even more telling against a view of object recognition that postulates the making of absolute judgments of fine quantitative detail is that the speed and accuracy of such judgments decline dramatically when they have to be made for multiple attributes (Egeth & Pachella, 1969; Garner, 1962; Miller, 1956). In contrast, object recognition latencies for complex objects are reduced by the presence of additional (redundant) components (Biederman, Ju, & Clapper, 1985, described below).

Geons Generated From Differences in Nonaccidental Properties Among Generalized Cones

I have emphasized the particular set of nonaccidental properties shown in Figure 4 because they may constitute a perceptual basis for the generation of the set of components. Any primitive that is hypothesized to be the basis of object recognition should be rapidly identifiable and invariant over viewpoint and noise. These characteristics would be attainable if differences among components were based on differences in nonaccidental properties. Although additional nonaccidental properties exist, there is empirical support for rapid perceptual access to the five described in Figure 4. In addition, these five relations reflect intuitions about significant perceptual and cognitive differences among objects.

From variation over only two or three levels in the nonaccidental relations of four attributes of generalized cylinders, a set of 36 geons can be generated. A subset is illustrated in Figure 6.

Six of the generated geons (and their attribute values) are shown in Figure 7. Three of the attributes describe characteristics of the cross section: its shape, symmetry, and constancy of size as it is swept along the axis. The fourth attribute describes the shape of the axis. Additional volumes are shown in Figures 8 and 9.

Nonaccidental Two-Dimensional Contrasts Among the Geons

As indicated in the above outline, the values of the four generalized cone attributes can be directly detected as contrastive differences in nonaccidental properties: straight versus curved, symmetrical versus asymmetrical, parallel versus nonparallel (and if nonparallel, whether there is a point of maximal convexity). Cross-section edges and curvature of the axis are distinguishable by collinearity or curvilinearity. The constant versus expanded size of the cross section would be detectable through parallelism; a constant cross section would produce a general-

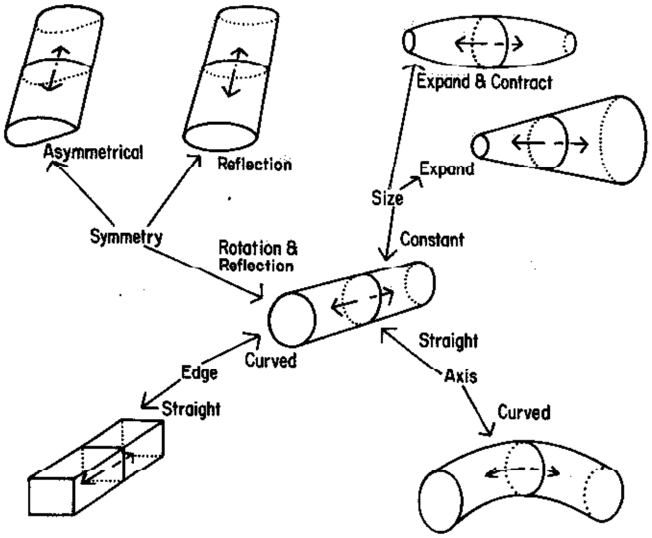


Figure 6. An illustration of how variations in three attributes of a cross section (curved vs. straight edges; constant vs. expanded vs. expanded and contracted size; mirror and rotational symmetry vs. mirror symmetry vs. asymmetrical) and one of the shape of the axis (straight vs. curved) can generate a set of generalized cones differing in nonaccidental relations. (Constant-sized cross sections have parallel sides; expanded or expanded and contracted cross sections have sides that are not parallel. Curved versus straight cross sections and axes are detectable through collinearity or curvature. The three values of cross-section symmetry [symmetrical under reflection & 90° rotation, reflection only, or asymmetrical] are detectable through the symmetry relation. Neighbors of a cylinder are shown here. The full family of geons has 36 members.)

ized cone with parallel sides (as with a cylinder or brick); an expanded cross section would produce edges that were not parallel (as with a cone or wedge). A cross section that expanded and then contracted would produce an ellipsoid with nonparallel sides and extrema of positive curvature (as with a lemon). Such extrema are invariant with viewpoint (e.g., Hoffman & Richards, 1985) and actually constitute a sixth nonaccidental relation. The three levels of cross-section symmetry are equivalent to Garner's (1974) distinction as to the number of different stimuli produced by increments of 90° rotations and reflections of a stimulus. Thus, a square or circle would be invariant under 90° rotation and reflection, but a rectangle or ellipse would be invariant only under reflection, as 90° rotations would produce another figure in each case. Asymmetrical figures would produce eight different figures under 90° rotation and reflection.

Specification of the nonaccidental properties of the three attributes of the cross section and one of the axis, as described in the previous paragraph, is sufficient to uniquely classify a given arrangement of edges as one of the 36 geons. These would be matched against a structural description for each geon that specified the values of these four nonaccidental image properties. But there are actually more distinctive nonaccidental image features for each geon than the four described in the previous paragraph (or indicated in Figures 7, 8, and 9). In particular, the arrangement of vertices, both of the silhouette and the presence of an interior Y vertex, and the presence of a discontinuous (third) edge along the axis (which produces the interior Y vertex) provide a richer description for each component than do the four properties of the generating function. This point can be readily appreciated by considering, as an example, some of the additional nonaccidental properties differentiating the brick from the cylinder in Figure 5. Each geon's structural description would thus include a larger number of contrastive image properties than the four that were directly related to the generating function.

Consideration of the featural basis for the structural descriptions for each geon suggests that a similarity measure can be defined on the basis of the common versus distinctive image features for any pair of components. The similarity measure would permit the promotion of alternative geons under conditions of ambiguity, as when one or several of the image features were undecidable.

Is geon identification two-dimensional or three-dimensional? Although the 36 geons have a clear subjective volumetric interpretation, it must be emphasized they can be uniquely specified from their two-dimensional image properties. Consequently, recognition need not follow the construction of an "object centered" (Marr, 1982) three-dimensional interpretation of each volume. It is also possible that, despite the subjective componential interpretation given to the arrangement of image features as simple volumes, it is the image features themselves, in specified relationships, that mediate perception. These alternatives remain to be evaluated.

Additional Sources of Contour and Recognition Variation

RBC seeks to account for the recognition of an infinitely varied perceptual input with a modest set of idealized primitives.

Partial Tentative Geon Set Based on Nonaccidentalness Relations

CROSS SECTION

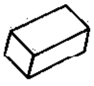
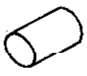
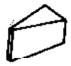
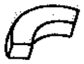
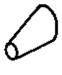
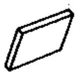
Geon	Edge Straight S Curved C	Symmetry Rot & Ref ++ Ref + Asymm –	Size Constant ++ Expanded – Exp & Cont ––	Axis Straight + Curved –
	S	++	++	+
	C	++	++	+
	S	+	–	+
	S	++	+	–
	C	++	–	+
	S	+	+	+

Figure 7. Proposed partial set of volumetric primitives (geons) derived from differences in nonaccidental properties.

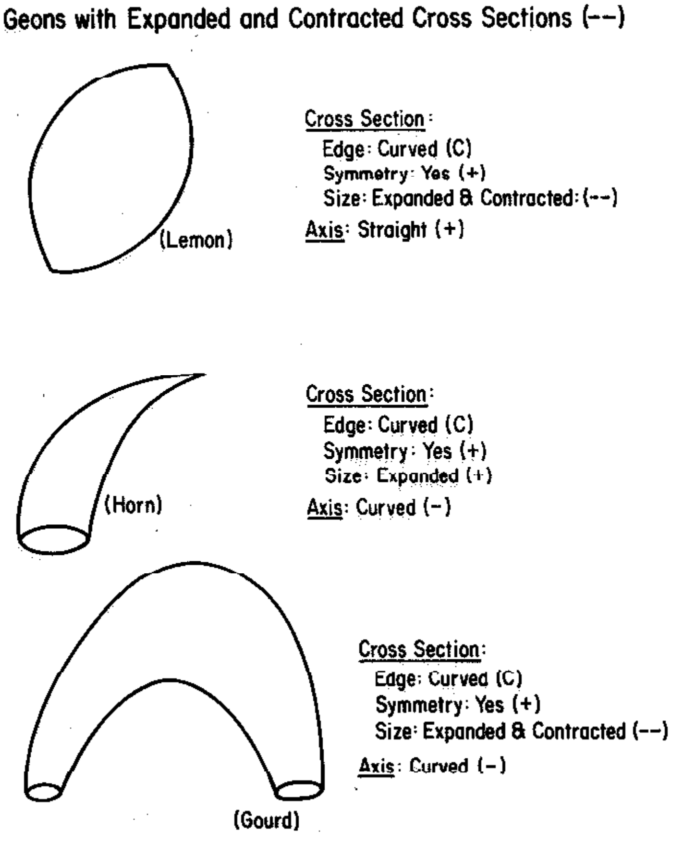


Figure 8. Three curved geons with curved axes or expanded and/or contracted cross sections. (These tend to resemble biological forms.)

A number of subordinate and related issues are raised by this attempt, some of which will be addressed in this section. This section need not be covered by a reader concerned primarily with the overall gist of RBC.

Asymmetrical cross sections. There are an infinity of possible cross sections that could be asymmetrical. How does RBC represent this variation? RBC assumes that the differences in the departures from symmetry are not readily available and thus do not affect primal access. For example, the difference in the shape of the cross section for the two straight-edged volumes in Figure 10 might not be apparent quickly enough to affect object recognition. This does not mean that an individual could not store the details of the volume produced by an asymmetrical cross section. But the presumption is that the access for this detail would be too slow to mediate primal access. I do not know of any case where primal access depends on discrimination among asymmetrical cross sections within a given component type, for example, among curved-edged cross sections of constant size, straight axes, and a specified aspect ratio. For instance, the curved cross section for the component that can model an airplane wing or car door is asymmetrical. Different wing designs might have different shaped cross sections. It is likely that most people, including wing designers, will know that the object is an airplane, or even an airplane wing, before they know the subclassification of the wing on the basis of the asymmetry of its cross section.

A second way in which asymmetrical cross sections need not be individually represented is that they often produce volumes that resemble symmetrical, but truncated, wedges or cones. This latter form of representing asymmetrical cross sections would be analogous to the schema-plus-correction phenomenon noted by Bartlett (1932). The implication of a schema-plus-correction representation would be that a single primitive category for asymmetrical cross sections and wedges might be sufficient. For both kinds of volumes, their similarity may be a function of the detection of a lack of parallelism in the volume. One would have to exert scrutiny to determine whether a lack of parallelism had originated in the cross section or in a size change of a symmetrical cross section. In this case, as with the components with curved axes described in the preceding section, a single primitive category for both wedges and asymmetrical straight-edged volumes could be postulated that would allow a reduction in the number of primitive components. There is considerable evidence that asymmetrical patterns require more time for their identification than symmetrical patterns (Checkosky & Whitlock, 1973; Pomerantz, 1978). Whether these effects have consequences for the time required for object identification is not yet known.

One other departure from regular components might also be noted. A volume can have a cross section with edges that are both curved and straight, as would result when a cylinder is sectioned in half along its length, producing a semicircular cross section. The conjecture is that in such cases the default cross section is the curved one, with the straight edges interpreted as slices off the curve, in schema-plus-correction representation (Bartlett, 1932).

CROSS SECTION

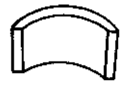
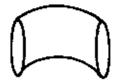
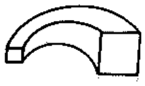
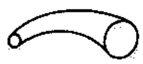
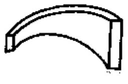
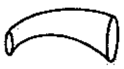
Geon	Edge Straight S Curved C	Symmetry Rot & Ref ++ Ref + Asymm −	Size Constant ++ Expanded − Exp & Cont −−	Axis Straight + Curved −
	S	+	++	−
	C	+	++	−
	S	++	−	−
	C	++	−	−
	S	+	−	−
	C	+	−	−

Figure 9. Geons with curved axis and straight or curved cross sections. (Determining the shape of the cross section, particularly if straight, might require attention.)

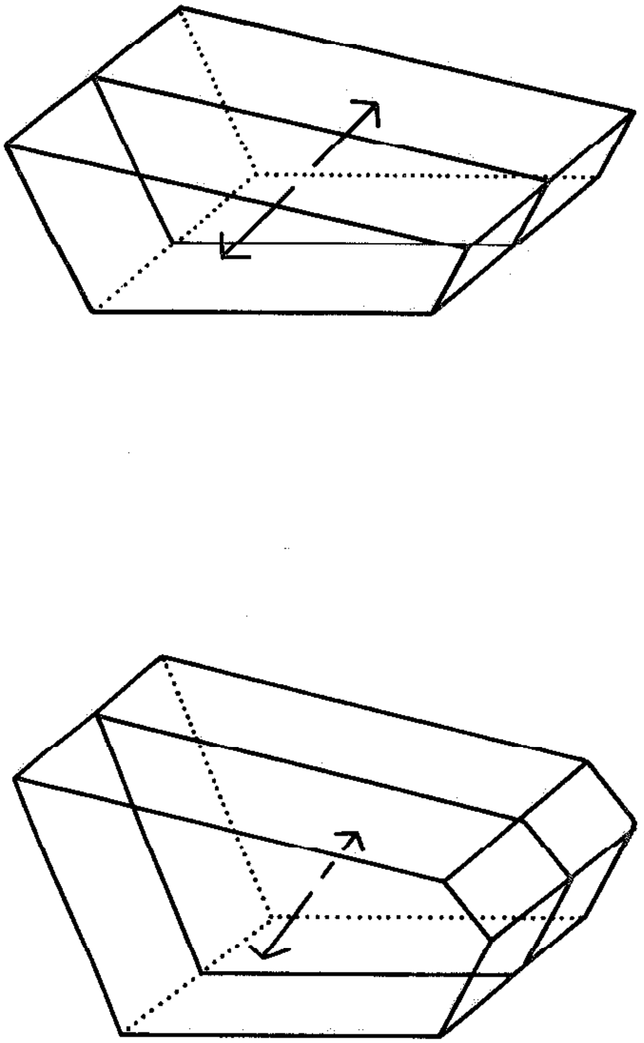


Figure 10. Volumes with an asymmetrical, straight-edged, cross section. (Detection of differences between such volumes might require attention.)

Component terminations. When a cross section varies in size, as with a cone, it can converge to a point, as with the horn in Figure 8, or appear truncated, as with the cone in Figure 7. Such termination differences could be represented as independently specified characteristics of the structural description for the geon, determinable in the image by whether the termination was a single L vertex (with a point) or two tangent Y vertices (with a truncated cone).

Another case arises when a cylinder has a cross section that remains constant for part of its length but then tapers to produce a point, as with a sharpened pencil. Such objects could be modeled by joining a cylinder to a cone, with the size of the cross sections matched so that no concavity is produced. The parsing point in this case would be the join where different nonaccidental properties were required to fit the volumes, namely, the shift from parallel edges with the cylinder to converging edges with the cone. Such joins provide a decidedly weaker basis—subjectively—for segmentation than joins producing cusps. The perceptual consequences of such variation have not been studied.

Metric variation. For any given geon type, there can be continuous metric variation in aspect ratio, degree of curvature (for curved components), and departure from parallelism (for nonparallel components). How should this quantitative variation be conceptualized? The discussion will concentrate on aspect ratio, probably the most important of the variations. But the issues will be generally applicable to the other metric variations as well.[6]

One possibility is to include specification of a range of aspect ratios in the structural description of the geons of an object as well as the object itself. It seems plausible to assume that recognition can be indexed, in part, by aspect ratio in addition to a componential description. An object's aspect ratio would thus play a role similar to that played by word length in the tachistoscopic identification of words, where long words are rarely proffered when a short word is flashed. Consider an elongated object, such as a baseball bat, with an aspect ratio of 15:1. When the orientation of the object is orthogonal to the viewpoint, so that the aspect ratio of its image is also 15:1, recognition might be faster than when presented at an orientation where the aspect ratio of its image differed greatly from that value, say 2:1. One need not have a particularly fine-tuned function for aspect ratio as large differences in aspect ratio between two components would, like parallelism, be preserved over a large proportion of arbitrary viewing angles.

Another way to incorporate variations in the aspect ratio of an object's image is to represent only qualitative differences, so that variations in aspect ratios exert an effect only when the relative size of the longest dimensions undergo reversal. Specifically, for each component and the complete object, three variations could be defined depending on whether the axis was much smaller, approximately equal to, or much longer than the longest dimension of the cross section. For example, for a geon whose axis was longer than the diameter of the cross section (which would be true in most cases), only when the projection of the cross section became longer than the axis would there be an effect of the object's orientation, as when the bat was viewed almost from on end so that the diameter of the handle was greater than the projection of its length.

A close dependence of object recognition performance on the preservation of the aspect ratio of a geon in the image would challenge RBC's emphasis on dichotomous contrasts of nonaccidental relations. Fortunately, these issues on the role of aspect ratio are readily testable. Bartram's (1976) experiments, described later in the section on orientation variability, suggest that sensitivity to variations in aspect ratio need not be given heavy weight: Recognition speed is unaffected by variation in aspect ratio across different views of the same object.

Planar geons. When a three-pronged vertex (viz., Y, tangent Y, or arrow) is not present in a parsed region, the resultant region appears planar, as with the flipper of the penguin in Figure

[6] Aspect ratio is a measure of the elongation of a component. For constant-sized cross sections and straight axes, it can be expressed as the width-to-height ratio of the smallest bounding rectangle that would just enclose the component. More complex functions are needed expressing the change in aspect ratio as a function of axis position when the cross section varies in size or the axis is curved.

10 or the eye of the elephant in Figure 11. Such shapes can be conceptualized in two ways. The first (and less favored) is to assume that these are just quantitative variations of the volumetric components, but with an axis length of zero. They would then have default values of a straight axis (+) and a constant cross section (+). Only the edge of the cross section and its symmetry could vary.

Alternatively, it might be that a planar region is not related perceptually to the foreshortened projection of the geon that could have produced it. Using the same variation in cross-section edge and symmetry as with the volumetric components, seven planar geons could be defined. For ++symmetry there would be the square and circle (with straight and curved edges, respectively) and for +symmetry the rectangle, triangle, and ellipse. Asymmetrical (−) planar geons would include trapezoids (straight edges), and drop shapes (curved edges). The addition of these seven planar geons to the 36 volumetric geons yields 43 components (a number close to the number of phonemes required to represent English words). The triangle is here assumed to define a separate geon, although a triangular cross section was not assumed to define a separate volume under the intuition that a prism (produced by a triangular cross section) is not quickly distinguishable from a wedge. My preference for assuming that planar geons are not perceptually related to their foreshortened volumes is based on the extraordinary difficulty of recognizing objects from views that are parallel to the axis of the major components so that foreshortening projects only the planar cross section, as shown in Figure 27. The presence of three-pronged vertices thus provides strong evidence that the image is generated from a volumetric rather than a planar component.

Selection of axis. Given that a volume is segmented from the object, how is an axis selected? Subjectively, it appears that an axis is selected that would maximize the axis's length, the symmetry of the cross section, and the constancy of the size of the cross section. By maximizing the length of the axis, bilateral symmetry can be more readily detected because the sides would be closer to the axis. Often a single axis satisfies all three criteria, but sometimes these criteria are in opposition and two (or more) axes (and component types) are plausible (Brady, 1983). Under such conditions, axes will often be aligned to an external frame, such as the vertical (Humphreys, 1983).

Negative values. The plus values in Figures 7, 8, and 9 are those favored by perceptual biases and memory errors. No bias is assumed for straight and curved edges of the cross section. For symmetry, clear biases have been documented. For example, if an image could have arisen from a symmetrical object, then it is interpreted as symmetrical (King et al., 1976). The same is apparently true of parallelism. If edges could be parallel, then they are typically interpreted as such, as with the trapezoidal room or window.

Curved axes. Figure 8 shows three of the most negatively marked primitives with curved crossed sections. Such geons often resemble biological entities. An expansion and contraction of a rounded cross section with a straight axis produces an ellipsoid (lemon), an expanded cross section with a curved axis produces a horn, and an expanded and contracted cross section with a rounded cross section produces a banana slug or gourd.

In contrast to the natural forms generated when both cross section and axis are curved, the geons swept by a straight-edged cross section traveling along a curved axis (e.g., the components on the first, third, and fifth rows of Figure 9) appear somewhat less familiar and more difficult to apprehend than their curved counterparts. It is possible that this difficulty may merely be a consequence of unfamiliarity. Alternatively, the subjective difficulty might be produced by a conjunction–attention effect (CAE) of the kind discussed by Treisman (e.g., Treisman & Gelade, 1980). (CAEs are described later in the section on attentional effects.) In the present case, given the presence in the image of curves and straight edges (for the rectilinear cross sections with curved axis), attention (or scrutiny) may be required to determine which kind of segment to assign to the axis and which to assign to the cross section. Curiously, the problem does not present itself when a curved cross section is run along a straight axis to produce a cylinder or cone. The issue as to the role of attention in determining geons would appear to be empirically tractable using the paradigms created by Treisman and her colleagues (Treisman, 1982; Treisman & Gelade, 1980).

Conjunction–attentional effects. The time required to detect a single feature is often independent of the number of distracting items in the visual field. For example, the time it takes to detect a blue shape (a square or a circle) among a field of green distractor shapes is unaffected by the number of green shapes. However, if the target is defined by a conjunction of features, for example, a blue square among distractors consisting of green squares and blue circles, so that both the color and the shape of each item must be determined to know if it is or is not the target, then target detection time increases linearly with the number of distractors (Treisman & Gelade, 1980). These results have led to a theory of visual attention that assumes that humans can monitor all potential display positions simultaneously and with unlimited capacity for a single feature (e.g., something blue or something curved). But when a target is defined by a conjunction of features, then a limited capacity attentional system that can only examine one display position at a time must be deployed (Treisman & Gelade, 1980).

The extent to which Treisman and Gelade's (1980) demonstration of conjunction–attention effects may be applicable to the perception of volumes and objects has yet to be evaluated. In the extreme, in a given moment of attention, it may be the case that the values of the four attributes of the components are detected as independent features. In cases where the attributes, taken independently, can define different volumes, as with the shape of cross sections and axes, an act of attention might be required to determine the specific component generating those attributes: Am I looking at a component with a curved cross section and a straight axis or is it a straight cross section and a curved axis? At the other extreme, it may be that an object recognition system has evolved to allow automatic determination of the geons.

The more general issue is whether relational structures for the primitive components are defined automatically or whether a limited attentional capacity is required to build them from their individual-edge attributes. It could be the case that some of the most positively marked geons are detected automatically, but that the volumes with negatively marked attributes might require attention. That some limited capacity is involved in the perception of objects (but not necessarily their components) is documented by an effect of the number of distracting objects on perceptual search (Biederman, Blickle, Teitelbaum, Klatsky,

& Mezzgnotte, in press). In their experiment, reaction times and errors for detecting an object such as a chair increased linearly as a function of the number of nontarget objects in a 100-ms presentation of nonscene arrangements of objects. Whether this effect arises from the necessity to use a limited capacity to construct a geon from its attributes or whether the effect arises from the matching of an arrangement of geons to a representation is not yet known.

Relations of RBC to Principles of Perceptual Organization

Textbook presentations of perception typically include a section of Gestalt organizational principles. This section is almost never linked to any other function of perception. RBC posits a specific role for these organizational phenomena in pattern recognition. As suggested by the section on generating geons through nonaccidental properties, the Gestalt principles, particularly those promoting Pragnanz (Good Figure), serve to determine the individual geons, rather than the complete object. A complete object, such as a chair, can be highly complex and asymmetrical, but the components will be simple volumes. A consequence of this interpretation is that it is the components that will be stable under noise or perturbation. If the components can be recovered and object perception is based on the components, then the object will be recognizable.

This may be the reason why it is difficult to camouflage objects by moderate doses of random occluding noise, as when a car is viewed behind foliage. According to RBC, the geons accessing the representation of an object can readily be recovered through routines of collinearity or curvature that restore contours (Lowe, 1984). These mechanisms for contour restoration will not bridge cusps (e.g., Kanizsa, 1979). For visual noise to be effective, by these considerations, it must obliterate the concavity and interrupt the contours from one geon at the precise point where they can be joined, through collinearity or constant curvature, with the contours of another geon. The likelihood of this occurring by moderate random noise is, of course, extraordinarily low, and it is a major reason why, according to RBC, objects are rarely rendered unidentifiable by noise. The consistency of RBC with this interpretation of perceptual organization should be noted. RBC holds that the (strong) loci of parsing is at cusps; the geons are organized from the contours between cusps. In classical Gestalt demonstrations, good figures are organized from the contours between cusps. Experiments subjecting these conjectures to test are described in a later section.

A Limited Number of Components?

According to the prior arguments, only 36 volumetric components can be readily discriminated on the basis of differences in nonaccidental properties among generalized cones. In addition, there are empirical and computational considerations that are compatible with a such a limit.

Empirically, people are not sensitive to continuous metric variations as evidenced by severe limitations in humans' capacity for making rapid and accurate absolute judgments of quantitative shape variations.[7] The errors made in the memory for shapes also document an insensitivity to metric variations. Computationally, a limit is suggested by estimates of the number of objects we might know and the capacity for RBC to readily represent a far greater number with a limited number of primitives.

Empirical Support for a Limit

Although the visual system is capable of discriminating extremely fine detail, I have been arguing that the number of volumetric primitives sufficient to model rapid human object recognition may be limited. It should be noted, however, that the number of proposed primitives is greater than the three—cylinder, sphere, and cone—advocated by some "How-to-Draw" books. Although these three may be sufficient for determining relative proportions of the parts of a figure and can aid perspective, they are not sufficient for the rapid identification of objects.[8] Similarly, Marr and Nishihara's (1978) pipe-cleaner (viz., cylinder) representations of animals (their Figure 17) would also appear to posit an insufficient number of primitives. On the page, in the context of other labeled pipe-cleaner animals, it is certainly possible to arrive at an identification of a particular (labeled) animal, for example, a giraffe. But the thesis proposed here would hold that the identifications of objects that were distinguished only by the aspect ratios of a single component type would require more time than if the representation of the object preserved its componential identity. In modeling only animals, it is likely that Marr and Nishihara capitalized on the possibility that appendages (such as legs and some necks) can often be modeled by the cylindrical forms of a pipe cleaner. By contrast, it is unlikely that a pipe-cleaner representation of a desk would have had any success. The lesson from Marr and Nishihara's demonstration, even when limited to animals, may be that an image that conveys only the axis structure and axes length is insufficient for primal access.

As noted earlier, one reason not to posit a representation system based on fine quantitative detail, for example, many variations in degree of curvature, is that such absolute judgments are notoriously slow and error prone unless limited to the 7 ± 2 values argued by Miller (1956). Even this modest limit is challenged when the judgments have to be executed over a brief 100-ms interval (Egeth & Pachella, 1969) that is sufficient for accurate object identification. A further reduction in the capacity for absolute judgments of quantitative variations of a simple

[7] Absolute judgments are judgments made against a standard in memory, for example, that Shape A is 14 cm. in length. Such judgments are to be distinguished from comparative judgments in which both stimuli are available for simultaneous comparison, for example, that Shape A, lying alongside Shape B, is longer than B. Comparative judgments appear limited only by the resolving power of the sensory system. Absolute judgments are limited, in addition, by memory for physical variation. That the memory limitations are severe is evidenced by the finding that comparative judgments can be made quickly and accurately for differences so fine that thousands of levels can be discriminated. But accurate absolute judgments rarely exceed 7 ± 2 categories (Miller, 1956).

[8] Paul Cezanne is often incorrectly cited on this point. "Treat nature by the cylinder, the sphere, the cone, *everything in proper perspective so that each side of an object or plane is directed towards a central point*" (Cezanne, 1904/1941, p. 234, italics mine). Cezanne was referring to perspective, not the veridical representation of objects.

shape would derive from the necessity, for most objects, to make simultaneous absolute judgments for the several shapes that constitute the object's parts (Egeth & Pachella, 1969; Miller, 1956). This limitation on our capacities for making absolute judgments of physical variation, when combined with the dependence of such variation on orientation and noise, makes quantitative shape judgments a most implausible basis for object recognition. RBC's alternative is that the perceptual discriminations required to determine the primitive components can be made categorically, requiring the discrimination of only two or three viewpoint-independent levels of variation.[9]

Our memory for irregular shapes shows clear biases toward "regularization" (e.g., Woodworth, 1938). Amply documented in the classical shape memory literature was the tendency for errors in the reproduction and recognition of irregular shapes to be in a direction of regularization, in which slight deviations from symmetrical or regular figures were omitted in attempts at reproduction. Alternatively, some irregularities were emphasized ("accentuation"), typically by the addition of a regular subpart. What is the significance of these memory biases? By the RBC hypothesis, these errors may have their origin in the mapping of the perceptual input onto a representational system based on regular primitives. The memory of a slight irregular form would be coded as the closest regularized neighbor of that form. If the irregularity was to be represented as well, an act that would presumably require additional time and capacity, then an additional code (sometimes a component) would be added, as with Bartlett's (1932) "schema with correction."

Computational Considerations: Are 36 Geons Sufficient?

Is there sufficient representational power in a set of 36 geons to express the human's capacity for basic-level visual categorizations? Two estimates are needed to provide a response to this question: (a) the number of readily available perceptual categories, and (b) the number of possible objects that could be represented by 36 geons. The number of possible objects that could be represented by 36 geons will depend on the allowable relations among the geons. Obviously, the value for (b) would have to be greater than the value for (a) if 36 geons are to prove sufficient.

How many readily distinguishable objects do people know? How might one arrive at a liberal estimate for this value? One estimate can be obtained from the lexicon. There are less than 1,500 relatively common basic-level object categories, such as chairs and elephants.[10] If we assume that this estimate is too small by a factor of 2, allowing for idiosyncratic categories and errors in the estimate, then we can assume potential classification into approximately 3,000 basic-level categories. RBC assumes that perception is based on a particular componential configuration rather than the basic-level category, so we need to estimate the mean number of readily distinguishable componential configurations per basic-level category. Almost all natural categories, such as elephants or giraffes, have one or only a few instances with differing componential descriptions. Dogs represent a rare exception for natural categories in that they have been bred to have considerable variation in their descriptions. Categories created by people vary in the number of allowable types, but this number often tends to be greater than the natural categories. Cups, typewriters, and lamps have just a few (in the case of cups) to perhaps 15 or more (in the case of lamps) readily discernible exemplars.[11] Let us assume (liberally) that the mean number of types is 10. This would yield an estimate of 30,000 readily discriminable objects (3,000 categories × 10 types/category).

A second source for the estimate derives from considering plausible rates for learning new objects. Thirty thousand objects would require learning an average of 4.5 objects per day, every day for 18 years, the modal age of the subjects in the experiments described below.

[9] This limitation on our capacities for absolute judgments also occurs in the auditory domain in speech perception, in which the modest number of phonemes can be interpreted as arising from dichotomous or trichotomous contrasts among a few invariant dimensions of speech production (Miller, 1956). Examples of invariant categorized speech features would be whether transitions are "feathered" (a cue for voicing) or the formants "murmured" (a cue for nasality). That these features are dichotomous allows the recognition system to avoid the limitations of absolute judgment in the auditory domain. It is possible that the limited number of phonemes derives more from this limitation for accessing memory for fine quantitative variation than it does from limits on the fineness of the commands to the speech musculature.

[10] This estimate was obtained from three sources: (a) several linguists and cognitive psychologists, who provided guesses of 300 to 1,000 concrete noun object categories; (b) the average 6-year-old child, who can name most of the objects seen in his or her world and on television and has a vocabulary of less than 10,000 words, about 10% of which are concrete count nouns; and (c) a 30-page sample from Webster's Seventh New Collegiate Dictionary, which provided perhaps the most defensible estimate; I counted the number of readily identifiable, unique concrete nouns that would not be subordinate to other nouns. Thus, "wood thrush" was not included because it could not be readily discriminated from "sparrow," but "penguin" and "ostrich" were counted as separate noun categories, as were borderline cases. The mean number of such nouns per page was 1.4, so given a 1,200 page dictionary, this is equivalent to 1,600 noun categories.

[11] It might be thought that faces constitute an obvious exception to the estimate of a ratio of ten exemplars per category presented here, in that we can obviously recognize thousands of faces. But can we recognize individual faces as rapidly as we recognize differences among basic level categories? I suspect not. That is, we may know that it is a face and not a chair in less time than that required for the identification of any particular face. Whatever the ultimate data on face recognition, there is evidence that the routines for processing faces have evolved to differentially respond to cuteness (Hildebrandt, 1982; Hildebrandt & Fitzgerald, 1983), age (e.g., Mark & Todd, 1985), and emotion and threats (e.g., Coss, 1979; Trivers, 1985). Faces may thus constitute a special stimulus case in that specific mechanisms have evolved to respond to biologically relevant quantitative variations and caution may be in order before results with face stimuli are considered characteristic of perception in general. Another possible exception to the exemplar/category ratio presented here occurs with categories such as lamps, which could have an arbitrarily large number of possible bases, shade types, and so on. But these variations may actually serve to hinder recognition. In a number of experiments in our laboratory, we have noted that highly stylized or unusual exemplars of a category are extremely difficult to identify under brief exposures (and out of context). The elements producing the variation in these cases may thus be acting as noise (or irrelevant components) in the sense that they are present in the image but not present in the mental representation for that category. These potential difficulties in the identification of faces or objects may not be subjectively apparent from the casual perusal of objects on a page, particularly when they are in a context that facilitates their classification.

Table 1
Generative Power of 36 Geons

Value	Component
36	First component (G_1)
×	×
36	Second component (G_2)
×	×
3	Size ($G_1 \gg G_2$, $G_1 \ll G_2$, $G_1 = G_2$)
×	×
2.4	G_1 top or bottom or side (represented for 80% of the objects)
×	×
2	Nature of join (end-to-end [off center] or end-to-side [centered])
×	×
2	Join at long or short surface of G_1
×	×
2	Join at long or short surface of G_2
	Total: 74,649 possible two-geon objects

Note. With three geons, 74,649 × 36 × 57.6 = 154 million possible objects. Equivalent to learning 23,439 new objects every day (approximately 1465/waking hr or 24/min) for 18 years.

Although the value of 4.5 objects learned per day seems reasonable for a child in that it approximates the maximum rates of word acquisition during the ages of 2–6 years (Carey, 1978), it certainly overestimates the rate at which adults develop new object categories. The impressive visual recognition competence of a 6-year-old child, if based on 30,000 visual categories, would require the learning of 13.5 objects per day, or about one per waking hour. By the criterion of learning rate, 30,000 categories would appear to be a liberal estimate.

Componential Relations: The Representational Capacity of 36 Geons

How many objects could be represented by 36 geons? This calculation is dependent upon two assumptions: (a) the number of geons needed, on average, to uniquely specify each object; and (b) the number of readily discriminable relations among the geons. We will start with (b) and see if it will lead to an empirically plausible value for (a). A possible set of relations is presented in Table 1. Like the components, the properties of the relations noted in Table 1 are nonaccidental in that they can be determined from virtually any viewpoint, are preserved in the two-dimensional image, and are categorical, requiring the discrimination of only two or three levels. The specification of these five relations is likely conservative because (a) it is certainly a nonexhaustive set in that other relations can be defined; and (b) the relations are only specified for a pair, rather than triples, of geons. Let us consider these relations in order of their appearance in Table 1.

1. Relative size. For any pair of geons, G_1 and G_2, G_1 could be much greater than, smaller than, or approximately equal to G_2.

2. Verticality. G_1 can be above or below or to the side of G_2, a relation, by the author's estimate, that is defined for at least 80% of the objects. Thus giraffes, chairs, and typewriters have a top-down specification of their components, but forks and knives do not. The handle of a cup is side-connected to the cylinder.

3. Centering. The connection between any pair of joined geons can be end-to-end (and of equal-sized cross section at the join), as the upper and lower arms of a person, or end-to-side, producing one or two concavities, respectively (Marr, 1977). Two-concavity joins are far more common in that it is rare that two arbitrarily joined end-to-end components will have equal-sized cross sections. A more general distinction might be whether the end of one geon in an end-to-side join is centered or off centered at the side of the other component. The end-to-end join might represent only the limiting, albeit special, case of off-centered joins. In general, the join of any two arbitrary volumes (or shapes) will produce two concavities, unless an edge from one volume is made to be joined and collinear with an edge from the other volume.

4. Relative size of surfaces at join. Other than the special cases of a sphere and a cube, all primitives will have at least a long and a short surface. The join can be on either surface. The attaché case in Figure 3A and the strongbox in Figure 3B differ by the relative lengths of the surfaces of the brick that are connected to the arch (handle). The handle on the shortest surface produces the strongbox; on a longer surface, the attaché case. Similarly, the cup and the pail in Figures 3C and 3D, respectively, differ as to whether the handle is connected to the long surface of the cylinder (to produce a cup) or the short surface (to produce a pail). In considering only two values for the relative size of the surface at the join, I am conservatively estimating the relational possibilities. Some volumes such as the wedge have as many as five surfaces, all of which can differ in size.

Representational Calculations

The 1,296 different pairs of the 36 geons (i.e., 36^2), when multiplied by the number of relational combinations, 57.6 (the product of the various values of the five relations), gives us 74,649 possible two-geon objects. If a third geon is added to the two, then this value has to be multiplied by 2,073 (36 geons × 57.6 ways in which the third geon can be related to one of the two geons), to yield 154 million possible three-component objects. This value, of course, readily accommodates the liberal estimate of 30,000 objects actually known.

The extraordinary disparity between the representational power of two or three geons and the number of objects in an individual's object vocabulary means that there is an extremely high degree of redundancy in the filling of the 154 million cell geon-relation space. Even with three times the number of objects estimated to be known by an individual (i.e., 90,000 objects), we would still have less than $\frac{1}{10}$ of 1% of the possible combinations of three geons actually used (i.e., over 99.9% redundancy).

There is a remarkable consequence of this redundancy if we assume that objects are distributed randomly throughout the object space. (Any function that yielded a relatively homogeneous distribution would serve as well.) The sparse, homogeneous occupation of the space means that, on average, it will be rare for an object to have a neighbor that differs only by one geon or relation.[12] Because the space was generated by considering only the number of possible two or three component objects, a constraint on the estimate of the average number of components per object that are sufficient for unambiguous identification is implicated. If objects were distributed relatively homogeneously among combinations of relations and geons, then only two or three geons would be sufficient to unambiguously represent most objects.

Experimental Support for a Componential Representation

According to the RBC hypothesis, the preferred input for accessing object recognition is that of the volumetric geons. In most cases, only a few appropriately arranged geons would be all that is required to uniquely specify an object. Rapid object recognition should then be possible. Neither the full complement of an object's geons, nor its texture, nor its color, nor the full bounding contour (or envelope or outline) of the object need be present for rapid identification. The problem of recognizing tens of thousands of possible objects becomes, in each case, just a simple task of identifying the arrangement of a few from a limited set of geons.

Several object-naming reaction time experiments have provided support for the general assumptions of the RBC hypothesis, although none have provided tests of the specific set of geons proposed by RBC or even that there might be a limit to the number of components.[13]

In all experiments, subjects named or quickly verified briefly presented pictures of common objects.[14] That RBC may provide a sufficient account of object recognition was supported by experiments indicating that objects drawn with only two or three of their components could be accurately identified from a single 100-ms exposure. When shown with a complete complement of components, these simple line drawings were identified almost as rapidly as full colored, detailed, textured slides of the same objects. That RBC may provide a necessary account of object recognition was supported by a demonstration that degradation (contour deletion), if applied at the regions that prevented recovery of the geons, rendered an object unidentifiable. All the original experimental results reported here have received at least one, and often several, replications.

Perceiving Incomplete Objects

Biederman, Ju, and Clapper (1985) studied the perception of briefly presented partial objects lacking some of their components. A prediction of RBC was that only two or three geons would be sufficient for rapid identification of most objects. If there was enough time to determine the geons and their relations, then object identification should be possible. Complete objects would be maximally similar to their representation and should enjoy an identification speed advantage over their partial versions.

Stimuli

The experimental objects were line drawings of 36 common objects, 9 of which are illustrated in Figure 11. The depiction of the objects and their partition into components was done subjectively, according to generally easy agreement among at least three judges. The artists were unaware of the set of geons described in this article. For the most part, the components corresponded to the parts of the object. Seventeen geon types (out of the full set of 36), were sufficient to represent the 180 components comprising the complete versions of the 36 objects.

The objects were shown either with their full complement of components or partially, but never with less than two components. The first two or three components that were selected were almost always the largest components from the complete object, as illustrated in Figures 12 and 13. For example, the airplane (Figure 13), which required nine components to look complete, had the fuselage and two wings when shown with three of its nine components. Additional components were added in decreasing order of size, subject to the constraint that additional components be connected to the existing components. Occasionally the ordering of large-to-small was altered when a smaller component, such as the eye of an animal, was judged to be highly diagnostic. The ordering by size was done under the assumption that processing would be completed earlier for larger components and, consequently, primal access would be controlled by those parts. However, it might be the case that a smaller part, if it was highly diagnostic, would have a greater role in controlling access than would be expected from its small size. The objects were displayed in black line on a white background and averaged 4.5° in greatest extent.

[12] Informal demonstrations suggest that this is the case. When a single component or relation of an object is altered, as with the cup and the pail, only with extreme rarity is a recognizable object from another category produced.

[13] Biederman (1985) discusses how a limit might be assessed. Among other consequences, a limit on the number of components would imply categorical effects whereby quantitative variations in the contours of an object, for example, degree of curvature, that did not alter a component's identity would have less of an effect on the identification of the object than contour variations that did alter a component's identity.

[14] Our decision to use a naming task with which to assess object recognition was motivated by several considerations. Naming is a sure sign of recognition. Under the conditions of these experiments, if an individual could name the object, he or she must have recognized it. With other paradigms, such as discrimination or verification, it is difficult (if not impossible) to prevent the subject from deriving stimulus selection strategies specific to the limited number of stimuli and distractors. Although naming RTs are relatively slow, they are remarkably well behaved, with surprisingly low variability (given their mean) for a given response and few of the response anticipation or selection errors that occur with binary responses (especially, keypresses). As in any task with a behavioral measure, one has to exert caution in making inferences about representations at an earlier stage. In every experiment reported here, whenever possible, the same objects (with the same name) served in all conditions. The data from these experiments (e.g., Figures 19 and 20) were so closely and reasonably associated with the contour manipulations as to preclude accounts based on a late name-selection stage. Moreover, providing the subjects with the set of possible names prior to an experiment, which might have been expected to affect response selection, had virtually no effect on performance. When objects could not be used as their own controls, as was necessary in studies of complexity, it was possible to experimentally or statistically control naming-stage variability because the determinants of this variability—specifically, name familiarity (which is highly correlated with frequency and age of acquisition) and length—are well understood.

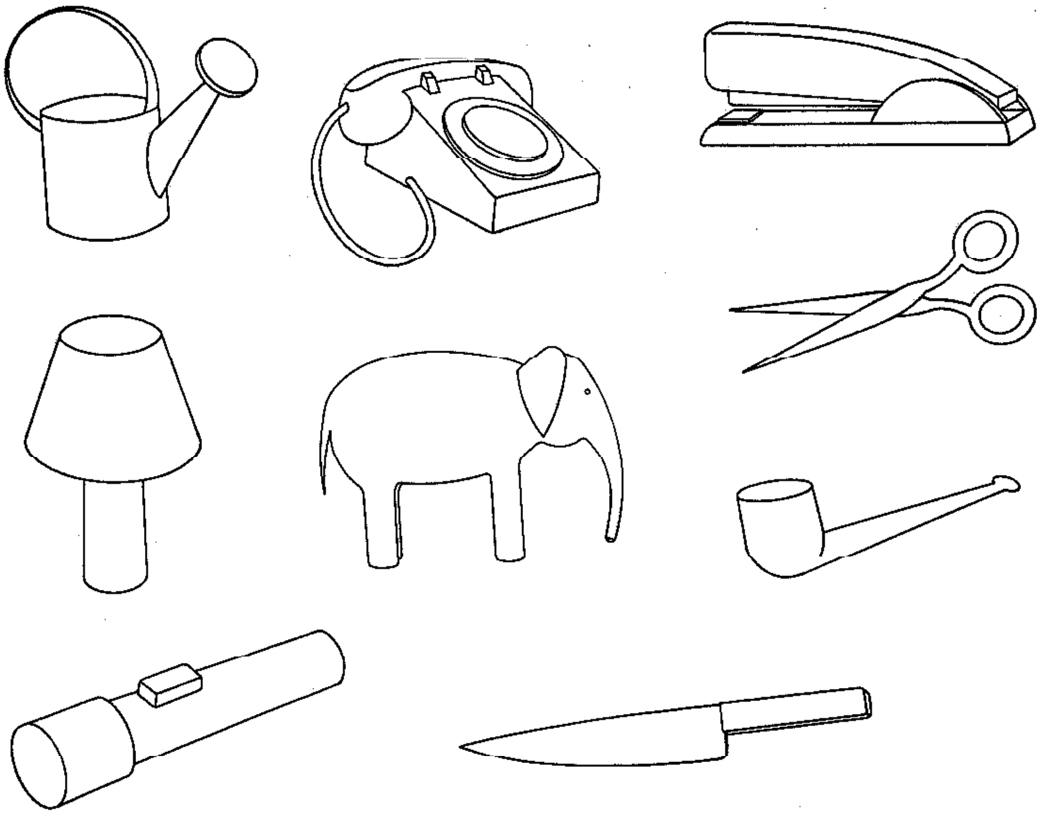


Figure 11. Nine of the experimental objects.

The purpose of this experiment was to determine whether the first few geons that would be available from an unoccluded view of a complete object would be sufficient for rapid identification of the object. We ordered the components by size and diagnosticity because our interest, as just noted, was on primal access in recognizing a complete object. Assuming that the largest and most diagnostic components would control this access, we studied the contribution of the *n*th largest and most diagnostic component, when added to the *n*−1 already existing components, because this would more closely mimic the contribution of that component when looking at the complete object. (Another kind of experiment might explore the contribution of an "average" component by balancing the ordering of the components. Such an experiment would be relevant to the recognition of an object that was occluded in such a way that only the displayed components would be available for viewing.)

Complexity

The objects shown in Figure 11 illustrate the second major variable in the experiment. Objects differ in complexity; by RBC's definition, the differences are evident in the number of components they require to look complete. For example, the lamp, the flashlight, the watering can, the scissors, and the elephant require two, three, four, six, and nine components, respectively. As noted previously, it would seem plausible that partial objects would require more time for their identification than complete objects, so that a complete airplane of nine components, for example, might be more rapidly recognized than only a partial version of that airplane, with only three of its components. The prediction from RBC was that complex objects, by furnishing more diagnostic combinations of components that could be simultaneously matched, would be more rapidly identified than simple objects. This prediction is contrary to models that assume that objects are recognized through a serial contour tracing process such as that studied by Ullman (1983).

General Procedure

Trials were self-paced. The depression of a key on the subject's terminal initiated a sequence of exposures from three projectors. First, the corners of a 500-ms fixation rectangle (6° wide) that corresponded to the corners of the object slide were shown. This fixation slide was immediately followed by a 100-ms exposure of a slide of an object that had varying numbers of its components present. The presentation of the object was immediately followed by a 500-ms pattern mask consisting of a random appearing arrangement of lines. The subject's task was to name the object as fast as possible into a microphone that triggered a voice key. The experimenter recorded errors. Prior to the experiment, the subjects read a list of the object names to be used in the experiment. (Subsequent experiments revealed that this procedure for name familiarization produced no effect. When subjects were not familiarized with the names of the experimental objects, results were virtually identical to

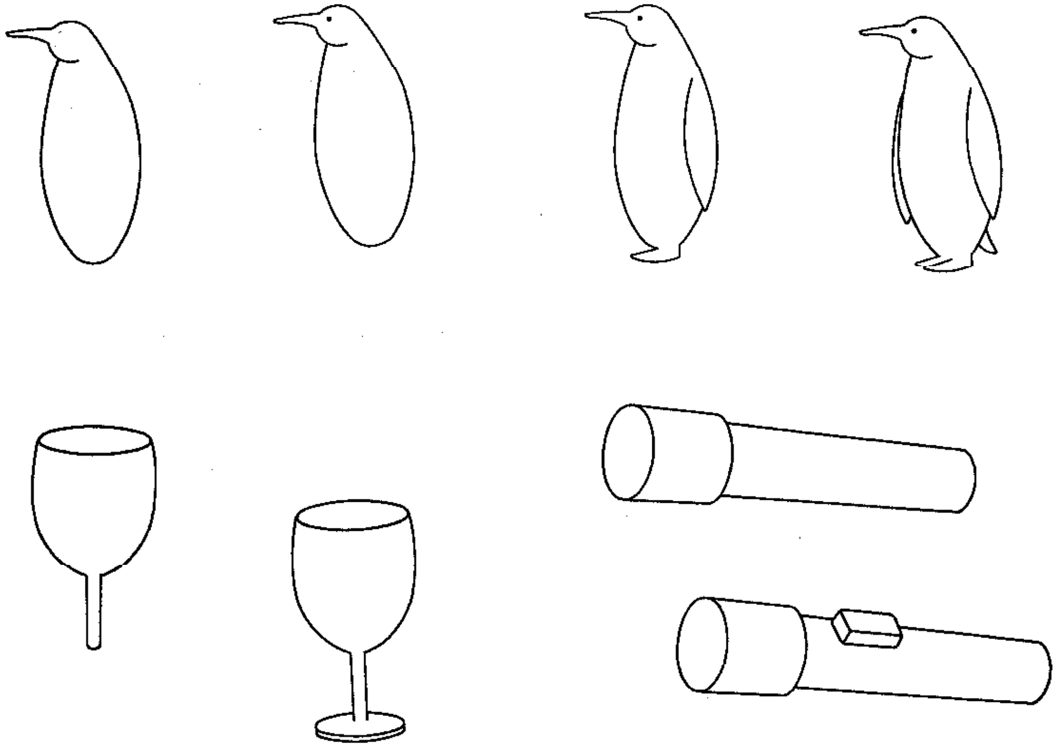


Figure 12. Illustration of the partial and complete versions of 2 three-component objects (the wine glass and flashlight) and 1 nine-component object (the penguin).

when such familiarization was provided. This finding indicates that the results of these experiments were not a function of inference over a small set of objects.) Even with the name familiarization, all responses that indicated that the object was identified were considered correct. Thus "pistol," "revolver," "gun," and "handgun" were all acceptable as correct responses for the same object. Reaction times (RTs) were recorded by a microcomputer that also controlled the projectors and provided speed and accuracy feedback on the subject's terminal after each trial.

Objects were selected that required two, three, six, or nine components to look complete. There were 9 objects for each of these complexity levels, yielding a total set of 36 objects. The various combinations of the partial versions of these objects brought the total number of experimental trials (slides) to 99. Each of 48 subjects viewed all the experimental slides, with balancing accomplished by varying the order of the slides.

Results

Figure 14 shows the mean error rates as a function of the number of components actually displayed on a given trial for the conditions in which no familiarization was provided. Each function is the mean for the nine objects at a given complexity level. Although each subject saw all 99 slides, only the data for the first time that a subject viewed a particular object will be discussed here. For a given level of complexity, increasing numbers of components resulted in better performance, but error rates overall were modest. When only three or four components of the complex objects (those with six or nine components to look complete) were present, subjects were almost 90% accurate. In general, the complete objects were named without error, so it is necessary to look at the RTs to see if differences emerge for the complexity variable.

Mean correct RTs, shown in Figure 15, provide the same general outcome as the errors, except that there was a slight tendency for the more complex objects, when complete, to have shorter RTs than the simple objects. This advantage for the complex objects was actually underestimated in that the complex objects had longer names (three and four syllables) and were less familiar than the simple objects. Oldfield (1966) and Oldfield and Wingfield (1965) showed that object-naming RTs were longer for names that have more syllables or are infrequent. This effect of slightly shorter RTs for naming complex objects has been replicated, and it seems safe to conclude, conservatively, that complex objects do not require more time for their identification than simple objects. This result is contrary to what would be expected from a serial contour-tracing process (e.g., Ullman, 1984). Serial tracing would predict that complex objects would require more time to be seen as complete compared to simple objects, which have less contour to trace. The slight RT advantage enjoyed by the complex objects is an effect that would be expected if their additional components were affording a redundancy gain from more possible diagnostic matches to their representations in memory.

Line Drawings Versus Colored Photography

The components that are postulated to be the critical units for recognition are edge-based and can be depicted by a line

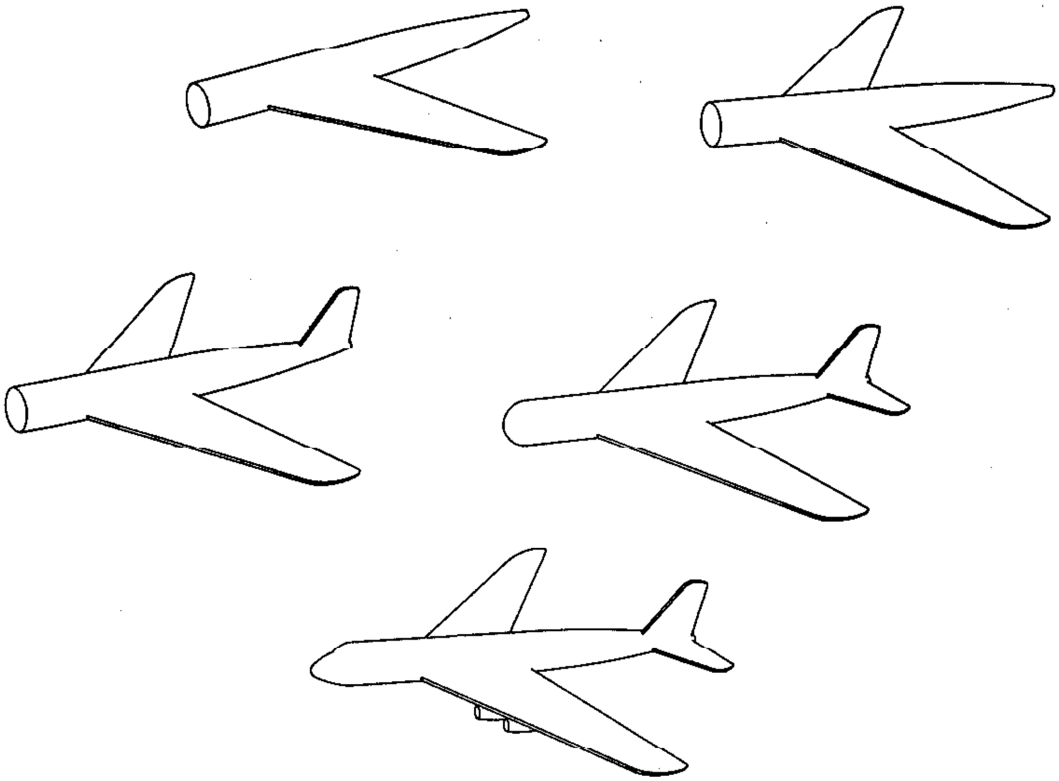


Figure 13. Illustration of partial and complete versions of a nine-component object (airplane).

drawing. Color, brightness, and texture would be secondary routes for recognition. From this perspective, Biederman and Ju (1986) reasoned that naming RTs for objects shown as line drawings should closely approximate naming RTs for those objects when shown as colored photographic slides with complete detail, color, and texture. This prediction would be true of any model that posited an edge-based representation mediating recognition.

In the Biederman and Ju experiments, subjects identified brief presentations (50–100 ms) of slides of common objects.[15] Each object was shown in two versions: professionally photographed in full color or as a simplified line drawing showing only the object's major components (such as those in Figure 11). In three experiments, subjects named the object; in a fourth experiment a yes–no verification task was performed against a target name. Overall, performance levels with the two types of stimuli were equivalent: mean latencies in identifying images presented by color photography were 11 ms shorter than the drawing but with a 3.9% higher error rate.

A previously unexplored color diagnosticity distinction among objects allowed us to determine whether color and lightness was providing a contribution to primal access independent of the main effect of photos versus drawings. For some kinds of objects, such as bananas, forks, fishes, or cameras, color is diagnostic to the object's identity. For other kinds, such as chairs, pens, or mittens, color is not diagnostic. The detection of a yellow region might facilitate the perception of a banana, but the detection of the color of a chair is unlikely to facilitate its identification, because chairs can be any color. If color was contributing to primal access, then the former kinds of objects, for which color is diagnostic, should have enjoyed a larger advantage when appearing in a color photograph, but this did not happen. Objects with a diagnostic color did not enjoy any advantage when they were displayed as color slides compared with their line-drawing versions. That is, showing color-diagnostic objects such as a banana or a fork as a color slide did not confer any advantage over the line-drawing version compared with objects such as a chair or mitten. Moreover, there was no color

[15] An oft-cited study, Ryan and Schwartz (1956), did compare photography (black & white) against line and shaded drawings and cartoons. But these investigators did not study basic-level categorization of an object. Subjects had to determine which one of four configurations of three objects (the positions of five double-throw electrical knife switches, the cycles of a steam valve, and the fingers of a hand) was being depicted. The subjects knew which object was to be presented on a given trial. For two of the three objects, the cartoons had lower thresholds than the other modes. But stimulus sampling and drawings and procedural specifications render interpretation of this experiment problematical; for example, the determination of the switch positions was facilitated in the cartoons by filling in the handles so they contrasted with the background contacts. The variability was enormous: Thresholds for a given form of depiction for a single object ranged across the four configurations from 50 to 2,000 ms. The cartoons did not have lower thresholds than the photographs for the hands, the stimulus example most frequently shown in secondary sources (e.g., Neisser, 1967; Hochberg, 1978; Rock, 1984). Even without a mask, threshold presentation durations were an order of magnitude longer than was required in the present study.

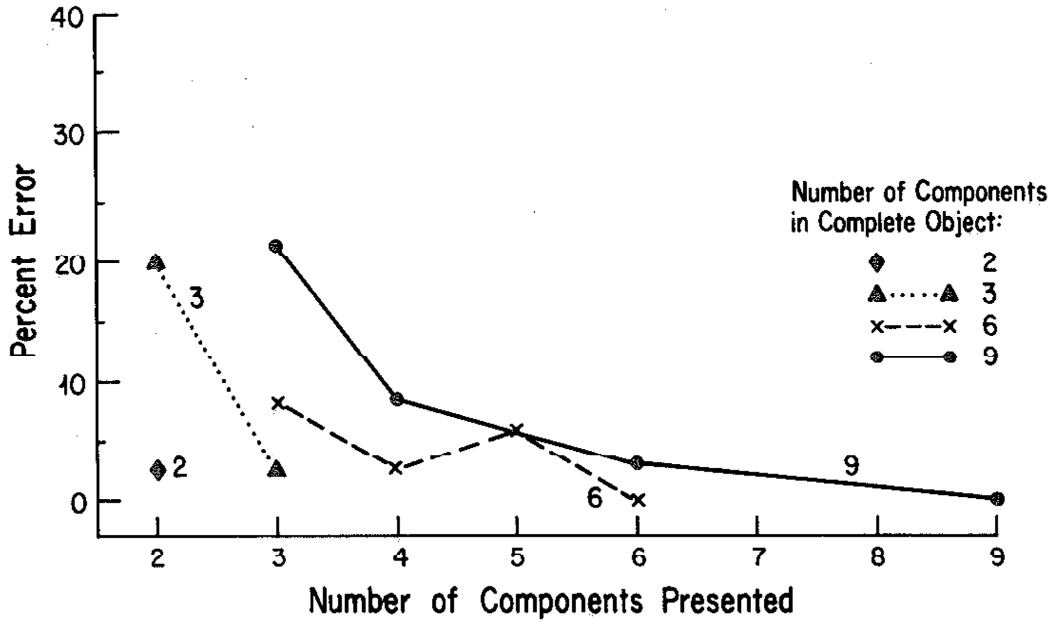


Figure 14. Mean percent error as a function of the number of components in the displayed object (abscissa) and the number of components required for the object to appear complete (parameter). (Each point is the mean for nine objects on the first occasion when a subject saw that particular object.)

diagnosticity advantage for the color slides on the verification task, where the color of the to-be-verified object could be anticipated.

This failure to find a color diagnosticity effect, when combined with the finding that simple line drawings could be identified so rapidly as to approach the naming speed of fully detailed, textured, colored photographic slides, supports the premise that the earliest access to a mental representation of an object can be modeled as a matching of an edge-based representation of a few simple components. Such edge-based descriptions are thus sufficient for primal access.

The preceding account should not be interpreted as suggesting that the perception of surface characteristics per se are delayed relative to the perception of the components but merely that in most cases surface cues are generally less efficient routes for primal access. That is, we may know that an image of a chair has a particular color and texture simultaneously with its volumetric description, but it is only the volumetric description that provides efficient access to the mental representation of "chair."

It should be noted that our failure to find a benefit from color photography is likely restricted to the domain whereby the edges are of high contrast. Under conditions where edge extraction is difficult, differences in color, texture, and luminance might readily facilitate such extraction and result in an advantage for color photography.

There is one surface characteristic that deserves special note: the luminance gradient. Such gradients can provide sufficient information as to a region's surface curvature (e.g., Besl & Jain, 1986) from which the surface's convexity or concavity can be determined. Our outline drawings lacked those gradients. Consider the cylinder and cone shown in the second and fifth rows, respectively, of Figure 7. In the absence of luminance gradients, the cylinder and cone are interpreted as convex (not hollow). Yet when the cylinder is used to make a cup and a pail in Figure 3, or the cone used to make a wine glass in Figure 12, the volumes are interpreted as concave (hollow). It would thus seem to be the case that the interpretation of hollowness—an interpretation that overrides the default value of solidity—of a volume can be readily accomplished top-down once a representation is elicited.

The Perception of Degraded Objects

RBC assumes that certain contours in the image are critical for object recognition. Several experiments on the perception of objects that have been degraded by deletion of their contour (Biederman & Blickle, 1985) provide evidence that these contours are necessary for object recognition (under conditions where contextual inference is not possible).

RBC holds that parsing of an object into components is performed at regions of concavity. The nonaccidental relations of collinearity and curvilinearity allow filling-in: They extend broken contours that are collinear or smoothly curvilinear. In concert, the two assumptions of (a) parsing at concavities and (b) filling-in through collinearity or smooth curvature lead to a prediction as to what should be a particularly disruptive form of degradation: If contours were deleted at regions of concavity in such a manner that their endpoints, when extended through collinearity or curvilinearity, bridge the concavity, then the components would be lost and recognition should be impossible. The cup in the right column of the top row of Figure 16 provides an example. The curve of the handle of the cup is drawn so that it is continuous with the curve of the cylinder forming the back rim of the cup. This form of degradation, in which the components cannot be recovered from the input through the nonaccidental properties, is referred to as *nonrecov-*

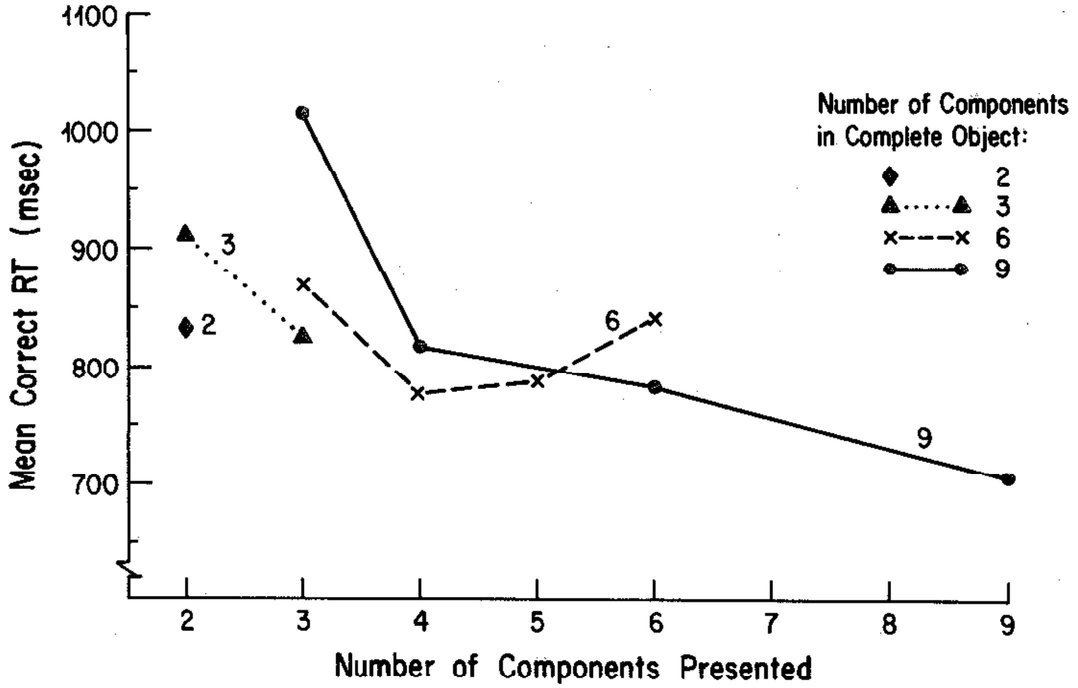


Figure 15. Mean correct reaction time as a function of the number of components in the displayed object (abscissa) and the number of components required for the object to appear complete (parameter). (Each point is the mean for nine objects on the first occasion when a subject saw that particular object.)

erable degradation and is illustrated for the objects in the right column of Figure 16.

An equivalent amount of deleted contour in a midsection of a curve or line should prove to be less disruptive as the components could then be restored through collinearity or curvature. In this case the components should be recoverable. Example of recoverable forms of degradation are shown in the middle column of Figure 16.

In addition to the procedure for deleting and bridging concavities, two other applications of nonaccidental properties were used to prevent determination of the components: vertex alteration and misleading symmetry or parallelism.

Vertex Alteration

When two or more edges terminate at the same point in the image, the visual system assumes that they are terminating at the same point in depth and a vertex is present at that point. Vertices are important for determining the nature of a component (see Figure 5). As noted previously, volumetric components will display at least one three-pronged vertex.

There are two ways to alter vertices. One way is by deleting a segment of an existing vertex. For example, the T-vertex produced by the occlusion of one blade of the scissors by the other has been converted into an L-vertex, suggesting that the boundaries of the region in the image are the boundaries of that region of the object. In the cup, the curved-T-vertex produced by the joining of a discontinuous edge of the front rim of the cup with the occlusional edge of the sides and back rim has been altered to an L-vertex by deleting the discontinuous edge. With only L-vertices, objects typically lose their volumetric character and appear planar.

The other way to alter vertices is to produce them through misleading extension of contours. Just as approximate joins of interrupted contours might be accepted to produce continuous edges, if three or more contours appear to meet at a common point when extended then a misleading vertex can be suggested. For example, in the watering can in the right column of Figure 11, the extensions of the contour from the spout attachment and sprinkler appear to meet the contours of the handle and rim, suggesting a false vertex of five edges. (Such a multivertex is nondiagnostic to a volume's three-dimensional identity [e.g., Guzman, 1968; Sugihara, 1984].)

Misleading Symmetry or Parallelism

Nonrecoverability of components can also be produced by contour deletion that produces symmetry or parallelism not characteristic of the original object. For example, the symmetry of oval region in the opening of the watering can suggests a planar component with that shape.

Even with these techniques, it was difficult to remove contours supporting all the components of an object, and some remained in nominally nonrecoverable versions, as with the handle of the scissors.

Subjects viewed 35 objects, in both recoverable and nonrecoverable versions. Prior to the experiment, all subjects were shown several examples of the various forms of degradation for several objects that were not used in the experiment. In addition, familiarization with the experimental objects was manipulated between subjects. Prior to the start of the experimental trials, different groups of six subjects (a) viewed a 3-sec slide of the intact version of the objects, for example, the objects in the left column of Figure 16, which they named; (b) were provided with the names of the objects on their terminal; or (c) were given no familiarization. As in the prior experiment, the subject's task was to name the objects.

A glance at the second and third columns in Figure 16 is sufficient to reveal that one does not need an experiment to show that the nonrecoverable objects would be more difficult to identify than the recoverable versions. But we wanted to determine if the nonrecoverable versions would be identifiable at extremely long exposure durations (5 s) and whether the prior exposure to the intact version of the object would overcome the effects of the contour deletion. The effects of contour deletion in the recoverable condition was also of considerable interest when compared with the comparable conditions from the partial object experiments.

Results

The error data are shown in Figure 17. Identifiability of the nonrecoverable stimuli was virtually impossible: The median error rate for those slides was 100%. Subjects rarely guessed wrong objects in this condition; most often they merely said that they "didn't know." When nonrecoverable objects could be identified, it was primarily for those instances where some of the components were not removed, as with the circular rings of the handle of the scissors. When this happened, subjects could name the object at 200-ms exposure duration. For the majority of the objects, however, error rates were well over 50% with no gain in performance even with 5 s of exposure duration. Objects in the recoverable condition were named at high accuracy at the longer exposure durations.

As in the previous experiments, familiarizing the subjects with the names of the objects had no effect compared with the condition in which the subjects were given no information about the objects. There was some benefit, however, in providing intact versions of the pictures of the objects. Even with this familiarity, performance in the nonrecoverable condition was extraordinarily poor, with error rates exceeding 60% when subjects had a full 5 s to decipher the stimulus. As noted previously, even this value underestimated the difficulty of identifying objects in the nonrecoverable condition, in that identification was possible only when the contour deletion allowed some of the components to remain recoverable.

The emphasis on the poor performance in the nonrecoverable condition should not obscure the extensive interference that was evident at the brief exposure durations in the recoverable condition. The previous experiments had established that intact objects, without picture familiarization, could be identified at near perfect accuracy at 100 ms. At this exposure duration in the present experiment, error rates for the recoverable stimuli, whose contours could be restored through collinearity and curvature, averaged 65%. These high error rates at 100-ms exposure duration suggest that the filling-in processes require an image (retinal or iconic)—not merely a memory representation—and sufficient time (on the order of 200 ms) to be successfully executed.

A Parametric Investigation of Contour Deletion

The dependence of componential recovery on the availability and locus of contour and time was explored parametrically by Biederman and Blickle (1985). In the previous experiment, it was necessary to delete or modify the vertices in order to produce the nonrecoverable versions of the objects. The recoverable versions of the objects tended to have their contours deleted in midsegment. It is possible that some of the interference in the nonrecoverable condition was a consequence of the removal of vertices per se, rather than the production of inappropriate components. Contour deletion was performed either at the vertices or at midsegments for 18 objects, but without the accidental bridging of components through collinearity or curvature that was characteristic of the nonrecoverable condition. The amount of contour removed varied from 25%, 45%, and 65%, and the objects were shown for 100, 200, or 750 ms. Other aspects of the procedure were identical to the previous experiments with only name familiarization provided. Figure 18 shows an example for a single object.

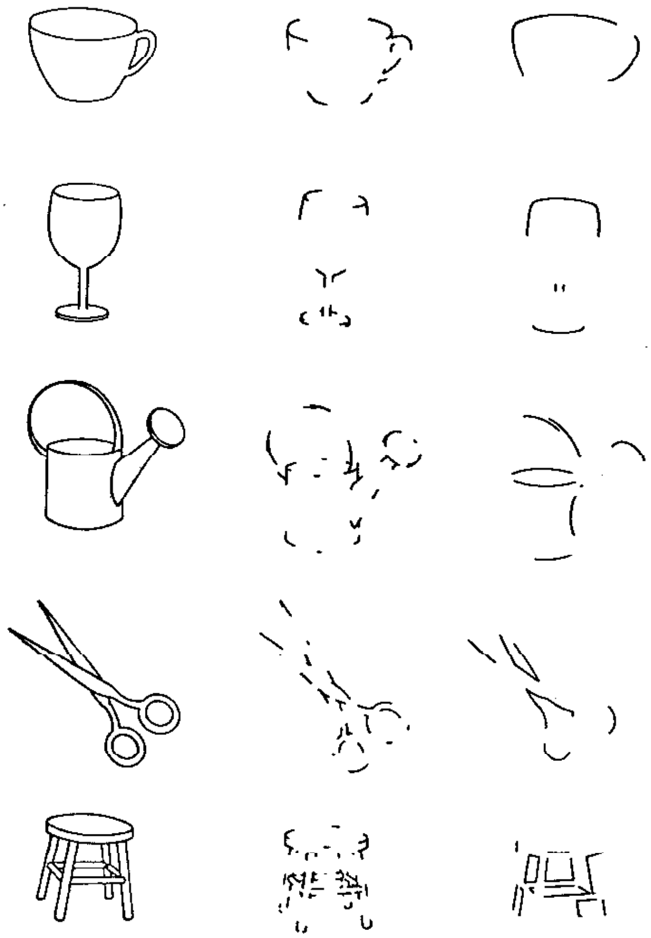


Figure 16. Example of five stimulus objects in the experiment on the perception of degraded objects. (The left column shows the original intact versions. The middle column shows the recoverable versions. The contours have been deleted in regions where they can be replaced through collinearity or smooth curvature. The right column shows the nonrecoverable versions. The contours have been deleted at regions of concavity so that collinearity or smooth curvature of the segments bridges the concavity. In addition, vertices have been altered, for example, from Ys to Ls, and misleading symmetry and parallelism have been introduced.)

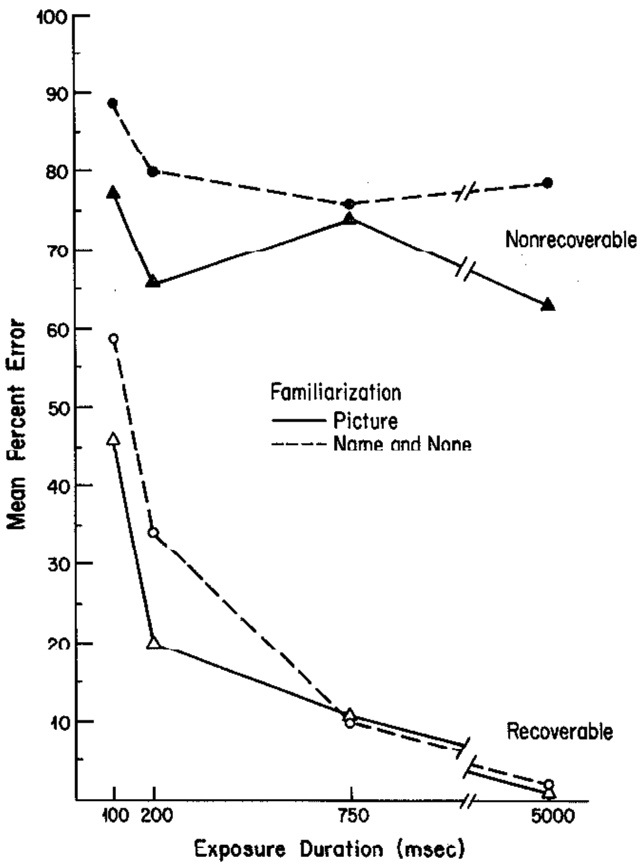


Figure 17. Mean percent errors in object naming as a function of exposure duration, nature of contour deletion (recoverable vs. nonrecoverable components), and familiarization (none, name, or picture). (No differences were apparent between the none and name pretraining conditions, so they have been combined into a single function.)

The mean percent errors are shown in Figure 19. At the briefest exposure duration and the most contour deletion (100-ms exposure duration and 65% contour deletion), removal of the vertices resulted in considerably higher error rates than the midsegment removal, 54% and 31% errors, respectively. With less contour deletion or longer exposures, the locus of the contour deletion had only a slight effect on naming accuracy. Both types of loci showed a consistent improvement with longer exposure durations, with error rates below 10% at the 750-ms duration. By contrast, the error rates in the nonrecoverable condition in the prior experiment exceeded 75%, even after 5 s. Although accuracy was less affected by the locus of the contour deletion at the longer exposure durations and the lower deletion proportions, there was a consistent advantage on naming latencies of the midsegment removal, as shown in Figure 20. (The lack of an effect at the 100-ms exposure duration with 65% deletion is likely a consequence of the high error rates for the vertex deletion stimuli.) This result shows that if contours are deleted at a vertex they can be restored, as long as there is no accidental filling-in. The greater disruption from vertex deletion is expected on the basis of their importance as diagnostic image features for the components. Overall, both the error and RT data document a striking dependence of object identification on what RBC assumes to be a prior and necessary stage of componential determination.

We conclude that the filling-in of contours, whether at midsegment or vertex, is a process that can be completed within 1 s. But the suggestion of a misleading component that bridges a concavity through collinearity or curvature produces an image that cannot index the original object, no matter how much time there is to view the image. Figure 21 compares a nonrecoverable version of an object (on the left) with a recoverable version, with considerably less contour available in the latter case. That the recoverable version is still identifiable shows that the recoverable objects would retain an advantage even if they had less contour than the nonrecoverable objects. Note that only four of the components in the recoverable version can be restored by the contours in the image, yet this is sufficient for recognition (although with considerable costs in time and effort). The recoverable version can be recognized despite the extreme distortion in the bounding contour and the loss of all the vertices from the right side of the object.

Perceiving Degraded Versus Partial Objects

Consider Figure 22 that shows, for some sample objects, one version in which whole components are deleted so that only three (of six or nine) of the components remain and another version in which the same amount of contour is removed, but in midsegment distributed over all of the object's components. With objects with whole components deleted, it is unlikely that the missing components are added imaginally, prior to recognition. Logically, one would have to know what object was being recognized to know what parts to add. Instead, the activation

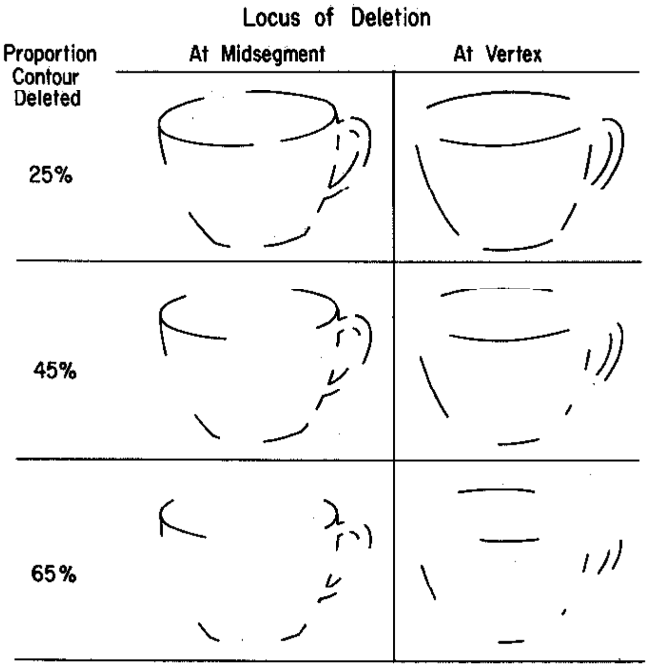


Figure 18. Illustration for a single object of 25, 45, and 65% contour removal centered at either midsegment or vertex. (Unlike the nonrecoverable objects illustrated in Figure 16, vertex deletion does not prevent identification of the object.)

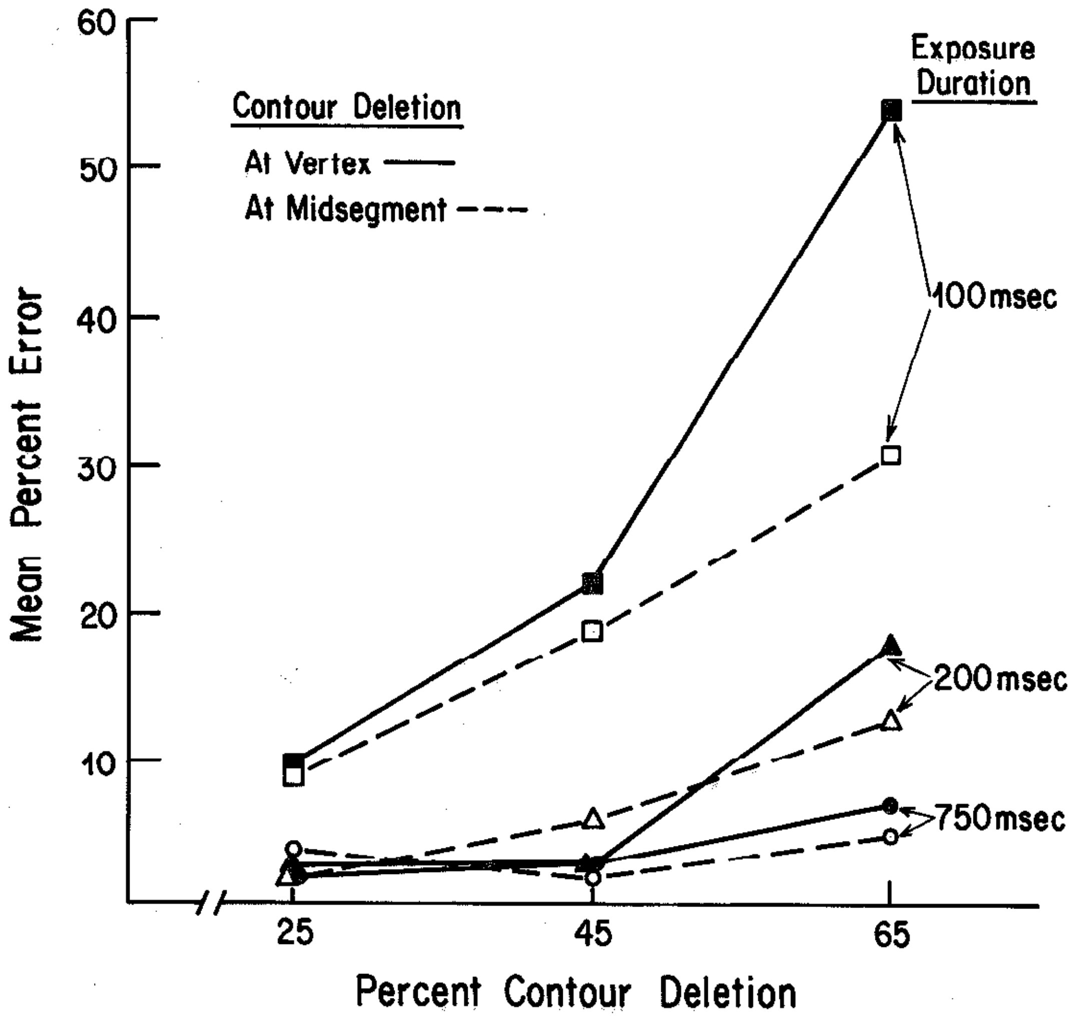


Figure 19. Mean percent object naming errors as a function of locus of contour removal (midsegment or vertex), percent removal, and exposure duration.

of a representation most likely proceeds in the absence of the parts, with weaker activation the consequence of the missing parts. The two methods for removing contour may thus be affecting different stages. Deleting contour in midsegment affects processes prior to and including those involved in the determination of the components (see Figure 2). The removal of whole components (the partial object procedure) is assumed to affect the matching stage, reducing the number of common components between the image and the representation and increasing the number of distinctive components in the representation. Contour filling-in is typically regarded as a fast, low-level process. We (Biederman, Beiring, Ju, & Blickle, 1985) studied the naming speed and accuracy of six- and nine-component objects undergoing these two types of contour deletion. At brief exposure durations (e.g., 65 ms) performance with partial objects was better than objects with the same amount of contour removed in midsegment both for errors (Figure 23) and RTs (Figure 24). At longer exposure durations (200 ms), the RTs reversed, with the midsegment deletion now faster than the partial objects.

Our interpretation of this result is that although a diagnostic subset of a few components (a partial object) can provide a sufficient input for recognition, the activation of that representation is not optimal compared with a complete object. Thus, in the partial object experiment described previously, recognition RTs were shortened with the addition of components to an already recognizable object. If all of an object's components were degraded (but recoverable), recognition would be delayed until contour restoration was completed. Once the filling-in was completed and the complete complement of an object's geons was activated, a better match to the object's representation would be possible (or the elicitation of its name) than with a partial object that had only a few of its components. The interaction can be modeled as a cascade in which the component-deletion condition results in more rapid activation of the geons but to a lower asymptote (because some geons never get activated) than the midsegment-deletion condition.

More generally, the finding that partial complex objects—with only three of their six or nine components present—can be recognized more readily than objects whose contours can be restored through filling-in documents the efficiency of a few components for accessing a representation.

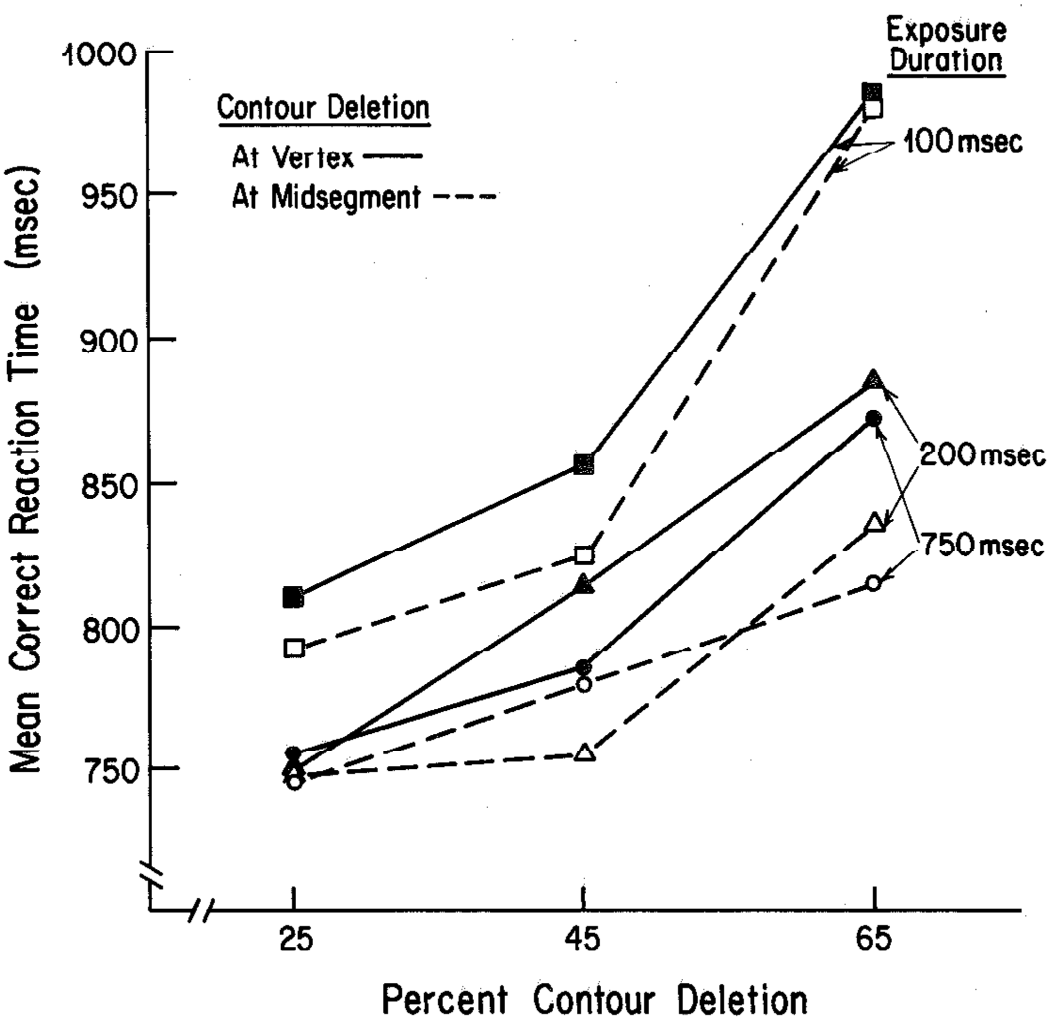


Figure 20. Mean correct object-naming reaction time (in milliseconds) as a function of locus of contour removal (midsegment or vertex), percent removal, and exposure duration.

Contour Deletion by Occlusion

The degraded recoverable objects in the right column of Figure 16 have the appearance of flat drawings of objects with interrupted contours. Biederman and Blickle (1985) designed a demonstration of the dependence of object recognition on componential identification by aligning an occluding surface so that it appeared to produce the deletions. If the components were responsible for an identifiable volumetric representation of the object, we would expect that with the recoverable stimuli the object would complete itself under the occluding surface and assume a three-dimensional character. This effect should not occur in the nonrecoverable condition. This expectation was met, as shown in Figures 25 and 26. These stimuli also provide a demonstration of the time (and effort?) requirements for contour restoration through collinearity or curvature. We have not yet obtained objective data on this effect, which may be complicated by masking effects from the presence of the occluding surface, but we invite the reader to share our subjective impressions. When looking at a nonrecoverable version of an object in Figure 25, no object becomes apparent. In the recoverable version in Figure 26, an object does pop into a three-dimensional appearance, but most observers report a delay (our own estimate is approximately 500 ms) from the moment the stimulus is first fixated to when it appears as an identifiable three-dimensional entity.

This demonstration of the effects of an occluding surface to produce contour interruption also provides a control for the possibility that the difficulty in the nonrecoverable condition was a consequence of inappropriate figure–ground groupings, as with the stool in Figure 16. With the stool, the ground that was apparent through the rungs of the stool became figure in the nonrecoverable condition. (In general, however, only a few of the objects had holes in them where this could have been a factor.) Figure–ground ambiguity would not invalidate the RBC hypothesis but would complicate the interpretation of the effects of the nonrecoverable noise, in that some of the effect would derive from inappropriate grouping of contours into components and some of the effect would derive from inappropriate figure–ground grouping. That the objects in the nonre-

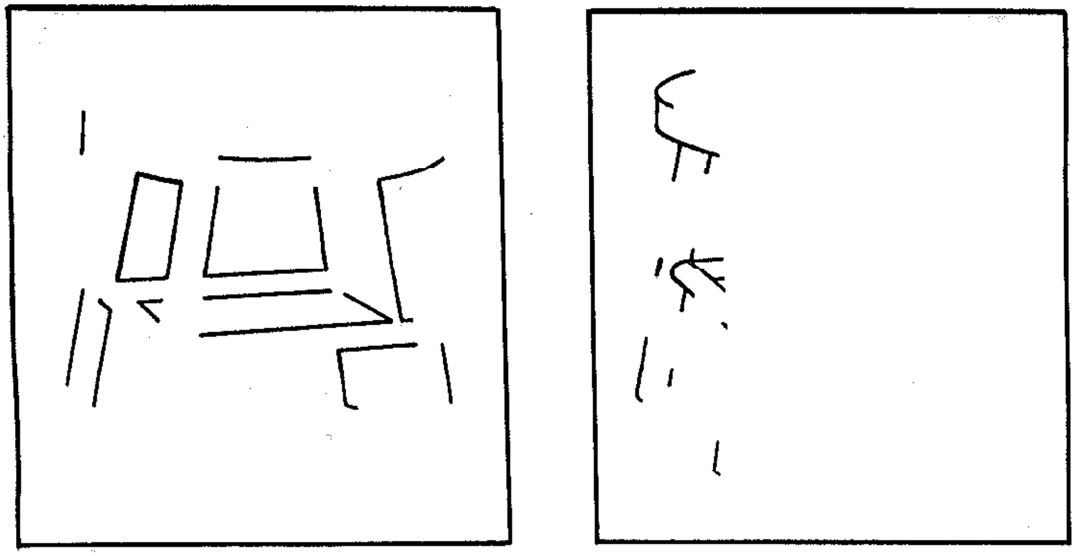


Figure 21. A comparison of a nonrecoverable version of an object (on the left) with a recoverable version (on the right) with half the contour of the nonrecoverable. Despite the reduction of contour the recoverable version still enjoys an advantage over the nonrecoverable.

coverable condition remain unidentifiable when the contour interruption is attributable to an occluding surface suggests that figure–ground grouping cannot be the primary cause of the interference from the nonrecoverable deletions.

Summary and Implications of the Experimental Results

The sufficiency of a component representation for primal access to the mental representation of an object was supported by two results: (a) that partial objects with two or three components could be readily identified under brief exposures, and (b) that line drawings and color photography produced comparable identification performance. The experiments with degraded stimuli established that the components are necessary for object perception. These results suggest an underlying principle by which objects are identified.

Principle of Componential Recovery

The results and phenomena associated with the effects of degradation and partial objects can be understood as the workings of a single Principle of Componential Recovery: If the components in their specified arrangement can be readily identified, object identification will be fast and accurate. In addition to those aspects of object perception for which experimental research was described above, the principle of componential recovery might encompass at least four additional phenomena in object perception: (a) objects can be more readily recognized from some orientations than from others (orientation variability); (b) objects can be recognized from orientations not previously experienced (object transfer); (c) articulated (or deformable) objects, with variable componential arrangements, can be recognized even when the specific configuration might not have been experienced previously (deformable object invariance); and (d) novel instances of a category can be rapidly classified (perceptual basis of basic-level categories).

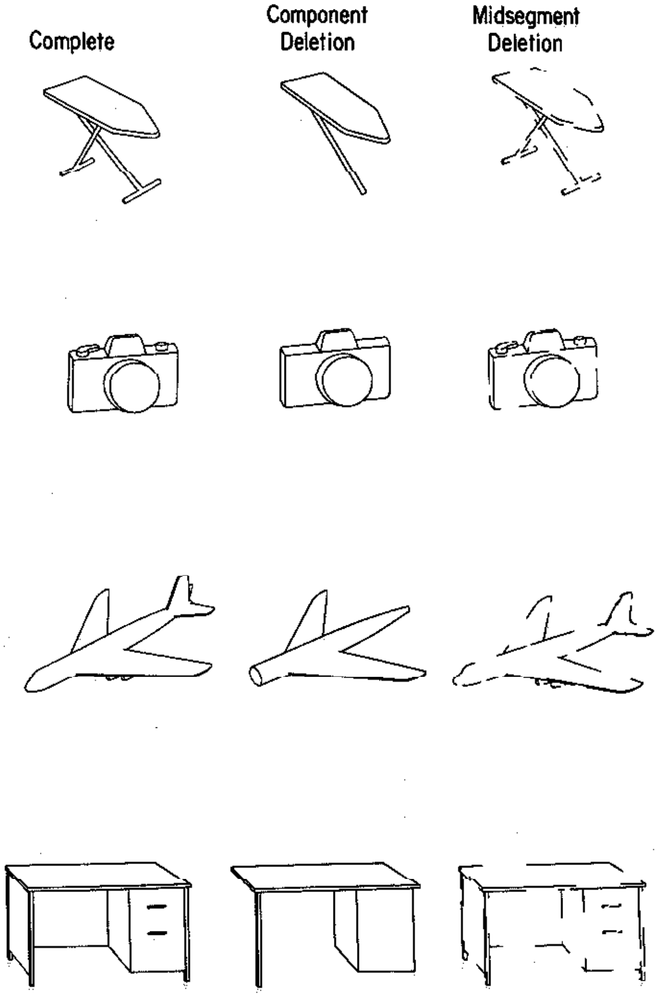


Figure 22. Sample stimuli with equivalent proportion of contours removed either at midsegments or as whole components.

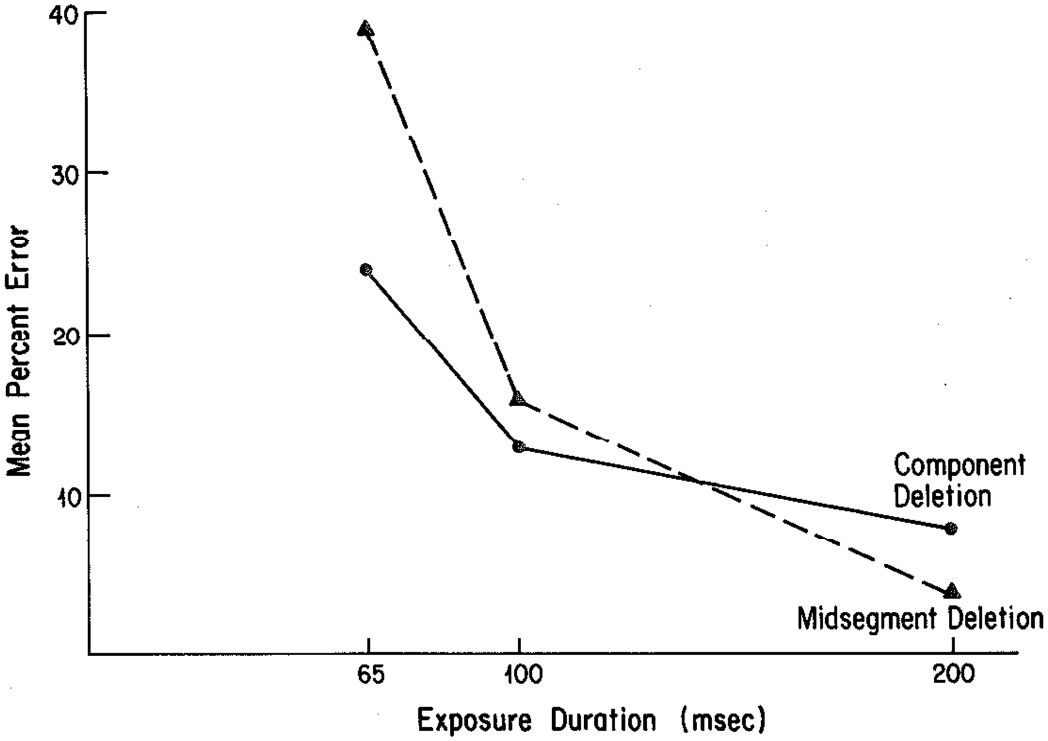


Figure 23. Mean percent errors of object naming as a function of the nature of contour removal (deletion of midsegments or components) and exposure duration.

Orientation Variability

Objects can be more readily identified from some orientations compared with others (Palmer, Rosch, & Chase, 1981). According to the RBC hypothesis, difficult views will be those in which the components extracted from the image are not the components (and their relations) in the representation of the object. Often such mismatches will arise from an "accident" of viewpoint where an image property is not correlated with the property in the three-dimensional world. For example, when the viewpoint in the image is along the axis of the major components of the object, the resultant foreshortening converts one or some of the components into surface components, such as disks and rectangles in Figure 27, which are not included in the componential description of the object. In addition, as illustrated in Figure 27, the surfaces may occlude otherwise diagnostic components. Consequently, the components extracted from the image will not readily match the mental representation of the object and identification will be much more difficult compared to an orientation, such as that shown in Figure 28, which does convey the components.

A second condition under which viewpoint affects identifiability of a specific object arises when the orientation is simply unfamiliar, as when a sofa is viewed from below or when the top–bottom relations among the components are perturbed as when a normally upright object is inverted. Jolicoeur (1985) recently reported that naming RTs were lengthened as a function of an object's rotation away from its normally upright position. He concluded that mental rotation was required for the identification of such objects, as the effect of X–Y rotation on RTs was similar for naming and mental rotation. It may be that mental rotation—or a more general imaginal transformation capacity stressing working memory—is required only under the (relatively rare) conditions where the relations among the components have to be rearranged. Thus, we might expect to find the equivalent of mental paper folding if the parts of an object were rearranged and the subject's task was to determine if a given object could be made out of the displayed components. RBC would hold that the lengthening of naming RTs in Jolicoeur's (1985) experiment is better interpreted as an effect that arises not from the use of orientation dependent features but from the perturbation of the "top-of" relations among the components.

Palmer et al. (1981) conducted an extensive study of the perceptibility of various objects when presented at a number of different orientations. Generally, a three-quarters front view was most effective for recognition, and their subjects showed a clear preference for such views. Palmer et al. (1981) termed this effective and preferred orientation of the object its *canonical orientation.* The canonical orientation would be, from the perspective of RBC, a special case of the orientation that would maximize the match of the components in the image to the representation of the object.

Transfer Between Different Viewpoints

When an object is seen at one viewpoint or orientation it can often be recognized as the same object when subsequently seen at some other orientation in depth, even though there can be extensive differences in the retinal projections of the two views. The principle of componential recovery would hold that trans-

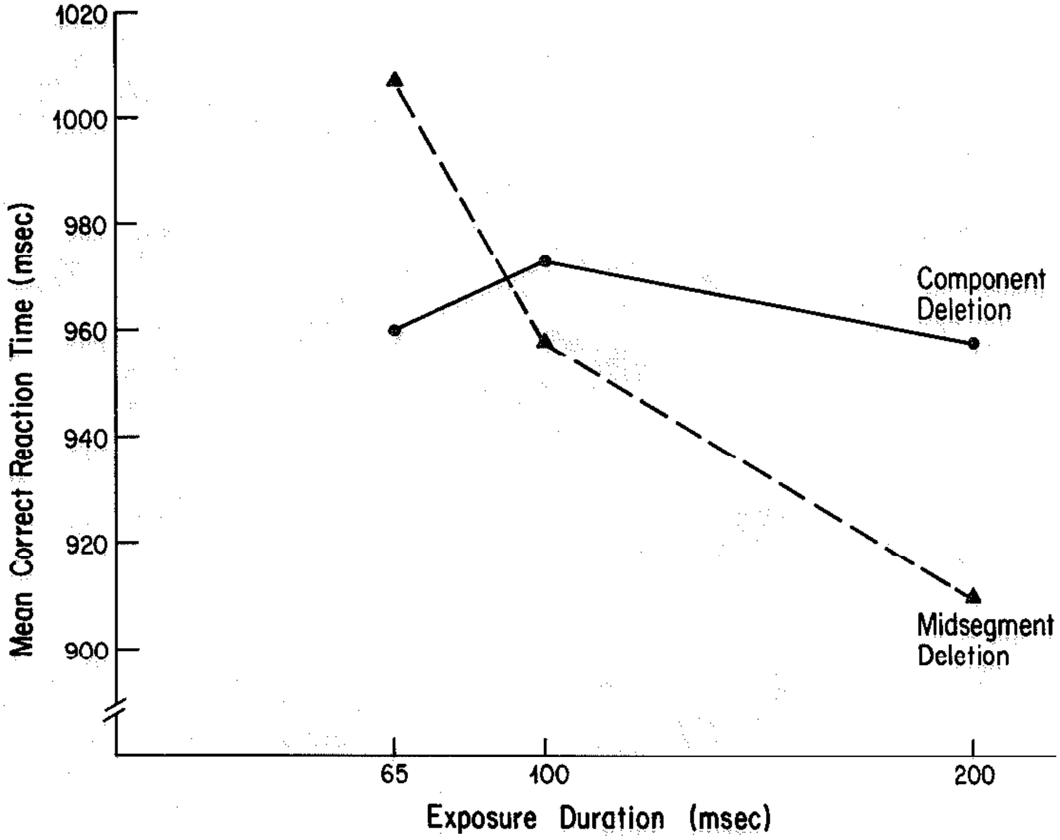


Figure 24. Mean correct reaction time (in milliseconds) in object naming as a function of the nature of contour removal (deletion at midsegments or components) and exposure duration.

fer between two viewpoints would be a function of the componential similarity between the views, as long as the relations among the components were not altered. This could be experimentally tested through priming studies with the degree of priming predicted to be a function of the similarity (viz., common minus distinctive components) of the two views. If two different views of an object contained the same components, RBC would predict that, aside from effects attributable to variations in aspect ratio, there should be as much priming as when the object was presented at an identical view. An alternative possibility to componential recovery is that a presented object would be mentally rotated (Shepard & Metzler, 1971) to correspond to the original representation. But mental rotation rates appear to be too slow and effortful to account for the ease and speed with which transfer occurs between different orientations in depth of the same object.

There may be a restriction on whether a similarity function for priming effects will be observed. Although unfamiliar objects (or nonsense objects) should reveal a componential similarity effect, the recognition of a familiar object, whatever its orientation, may be too rapid to allow an appreciable experimental priming effect. Such objects may have a representation for each orientation that provides a different componential description. Bartram's (1974) results support this expectation that priming effects might not be found across different views of familiar objects. Bartram performed a series of studies in which subjects named 20 pictures of objects over eight blocks of trials. (In another experiment [Bartram, 1976], essentially the same results were found with a same–different name-matching task in which pairs of pictures were presented.) In the *identical* condition, the pictures were identical across the trial blocks. In the *different view* condition, the same objects were depicted from one block to the next but in different orientations. In the *different exemplar* condition, different exemplars, for example, different instances of a chair, were presented, all of which required the same response. Bartram found that the naming RTs for the identical and different view conditions were equivalent and both were shorter than control conditions, described below, for concept and response priming effects. Bartram theorized that observers automatically compute and access all possible three-dimensional viewpoints when viewing a given object. Alternatively, it is possible that there was high componential similarity across the different views and the experiment was insufficiently sensitive to detect slight differences from one viewpoint to another. However, in four experiments with colored slides, we (Biederman & Lloyd, 1985) failed to obtain any effect of variation in viewing angle and have thus replicated Bartram's basic effect (or lack of effect). At this point, our inclination is to agree with Bartram's interpretation, with somewhat different language, but restrict its scope to familiar objects. It should be noted that both Bartram's and our results are inconsistent with a model that assigned heavy weight to the aspect ratio of the image of the object or postulated an underlying mental rotation function.

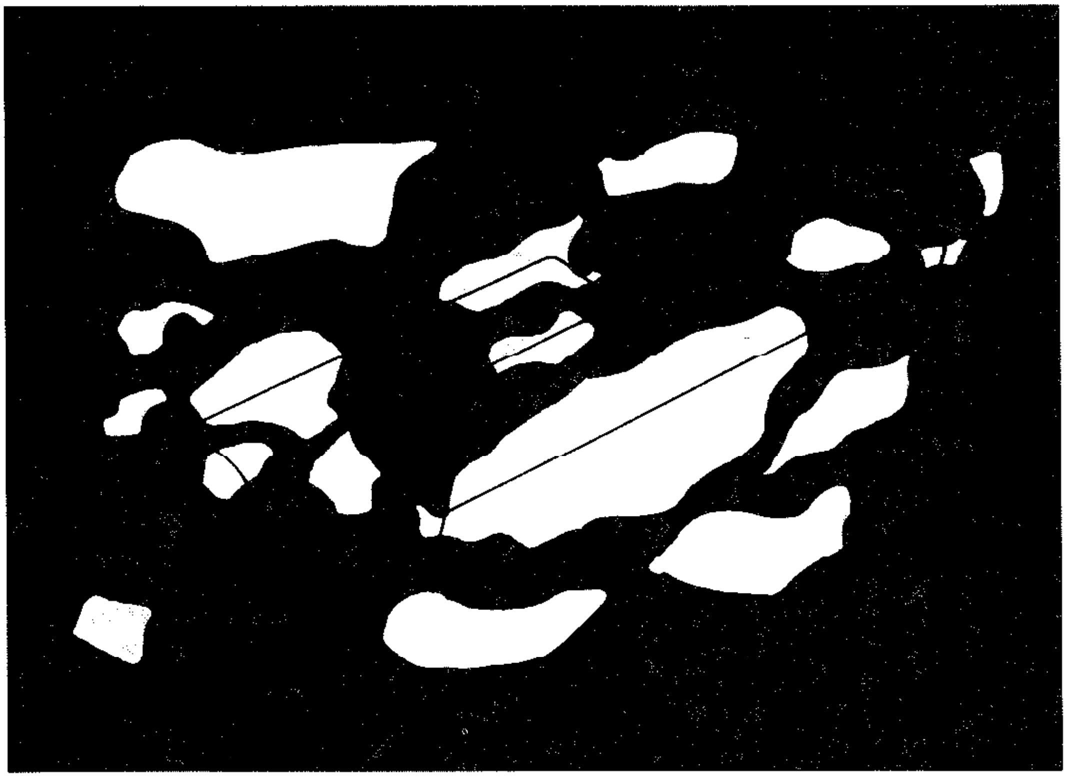


Figure 25. Nonrecoverable version of an object where the contour deletion is produced by an occluding surface.

Different Exemplars Within an Object Class

Just as we might be able to gauge the transfer between two different views of the same object based on a componential-based similarity metric, we might be able to predict transfer between different exemplars of a common object, such as two different instances of a lamp or chair.

As noted in the previous section, Bartram (1974) also included a different exemplar condition, in which different objects with the same name—different cars, for example—were depicted from block to block. Under the assumption that different exemplars would be less likely to have common components, RBC would predict that this condition would be slower than the identical and different view conditions but faster than a different object control condition with a new set of objects that required different names for every trial block. This was confirmed by Bartram.

For both different views of the same object as well as different exemplars (subordinates) within a basic-level category, RBC predicts that transfer would be based on the overlap in the components between the two views. The strong prediction would be that the same similarity function that predicted transfer between different orientations of the same object would also predict the transfer between different exemplars with the same name.

The Perceptual Basis of Basic Level Categories

Consideration of the similarity relations among different exemplars with the same name raises the issue as to whether objects are most readily identified at a basic, as opposed to a subordinate or superordinate, level of description. The componential representations described here are representations of specific, subordinate objects, although their identification was often measured with a basic-level name. Much of the research suggesting that objects are recognized at a basic level have used stimuli, often natural, in which the subordinate-level exemplars had componential descriptions that were highly similar to those for a basic-level prototype for that class of objects. Only small componential differences, or color or texture, distinguished the subordinate-level objects. Thus distinguishing Asian elephants from African elephants or Buicks from Oldsmobiles requires fine discrimination for their verification. The structural descriptions for the largest components would be identical. It is not at all surprising that in these cases basic-level identification would be most rapid. On the other hand, many human-made categories, such as lamps, or some natural categories, such as dogs (which have been bred by humans), have members that have componential descriptions that differ considerably from one exemplar to another, as with a pole lamp versus a ginger jar table lamp, for example. The same is true of objects that differ from their basic-level prototype, as penguins or sport cars. With such instances, which unconfound the similarity between basic-level and subordinate-level objects, perceptual access should be at the subordinate (or instance) level, a result supported by a recent report by Jolicoeur, Gluck, and Kosslyn (1984). In general, then, recognition will be at the subordinate level but will appear to be at the basic level when the componential descrip-

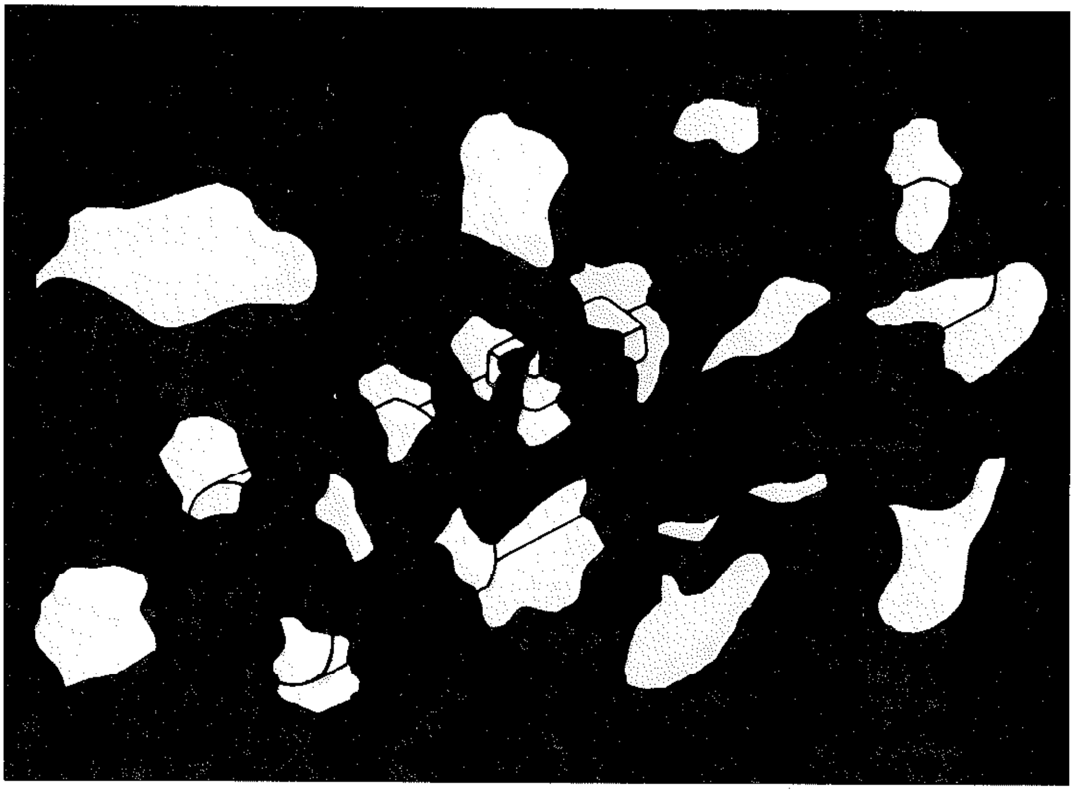


Figure 26. Recoverable version of an object where the contour deletion is produced by an occluding surface. (The object, a flashlight, is the same as that shown in Figure 25. The reader may note that the three-dimensional percept in this figure does not occur instantaneously.)

tions are the same at the two levels. However, the ease of perceptual recognition of nonprototypical exemplars, such as penguins, makes it clear that recognition will be at the level of the exemplar.

The kinds of descriptions postulated by RBC may play a central role in children's capacity to acquire names for objects. They may be predisposed to employ different labels for objects that have different geon descriptions. When the perceptual system presents a new description for an arrangement of large geons, the absence of activation might readily result in the question "What's that?"

For some categories, such as chairs, one can conceive of an extraordinarily large number of instances. Do we have a priori structural descriptions for all these cases? Obviously not. Although we can recognize many visual configurations as chairs, it is likely that only those for which there exists a close structural description in memory will recognition be rapid. The same caveat that was raised about the Marr and Nishihara (1978) demonstrations of pipe-cleaner animals in an earlier section must be voiced here. With casual viewing, particularly when supported by a scene context or when embedded in an array of other chairs, it is often possible to identify unusual instances as chairs without much subjective difficulty. But when presented as an isolated object without benefit of such contextual support, we have found that recognition of unfamiliar exemplars requires markedly longer exposure durations than those required for familiar instances of a category.

It takes but a modest extension of the principle of componential recovery to handle the similarity of objects. Simply put, similar objects will be those that have a high degree of overlap in their components and in the relations among these components. A similarity measure reflecting common and distinctive components (Tversky, 1977) may be adequate for describing the similarity among a pair of objects or between a given instance and its stored or expected representation, whatever their basic- or subordinate-level designation.

The Perception of Nonrigid Objects

Many objects and creatures, such as people and telephones, have articulated joints that allow extension, rotation, and even separation of their components. There are two ways in which such objects can be accommodated by RBC. One possibility, as described in the previous section on the representation for variation within a basic-level category, is that independent structural descriptions are necessary for each sizable alteration in the arrangement of an object's components. For example, it may be necessary to establish a different structural description for the left-most pose in Figure 29 than in the right-most pose. If this was the case, then a priming paradigm might not reveal any priming between the two stimuli. Another possibility is that the relations among the components can include a range of possible values (Marr & Nishihara, 1978). For a relation that allowed complete freedom for movement, the relation might simply be "joined." Even that might be relaxed in the case of objects with separable parts, as with the handset and base of a

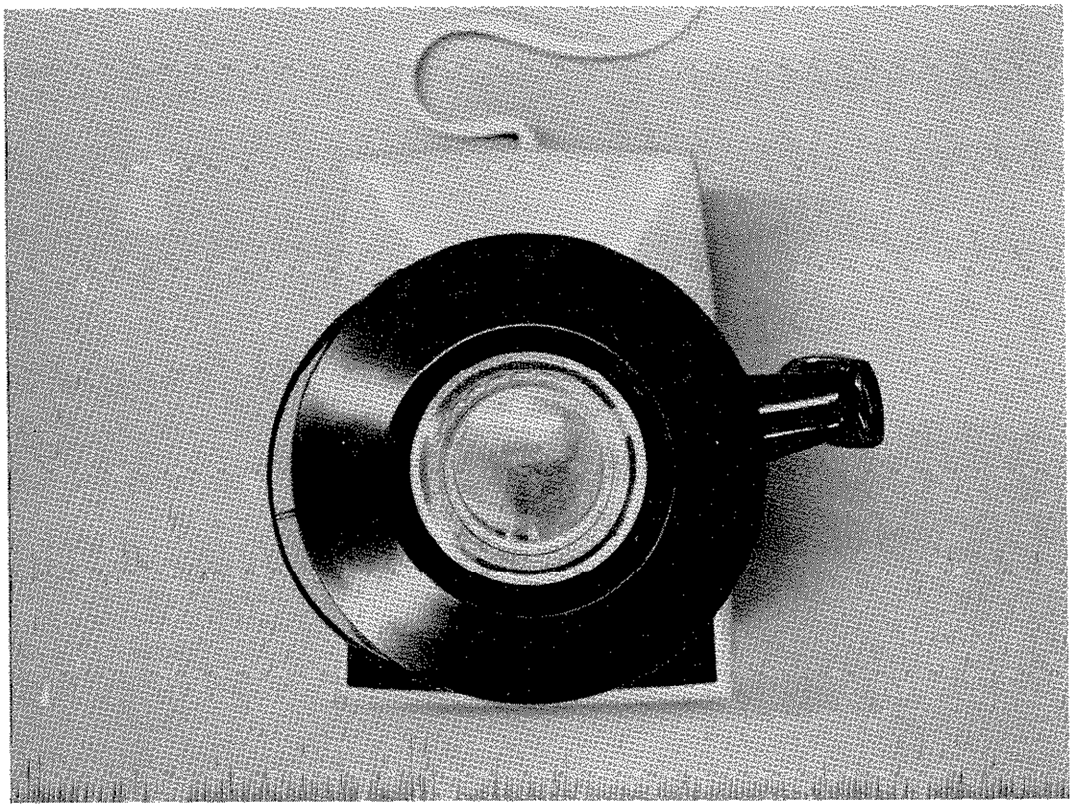


Figure 27. A viewpoint parallel to the axes of the major components of a common object.

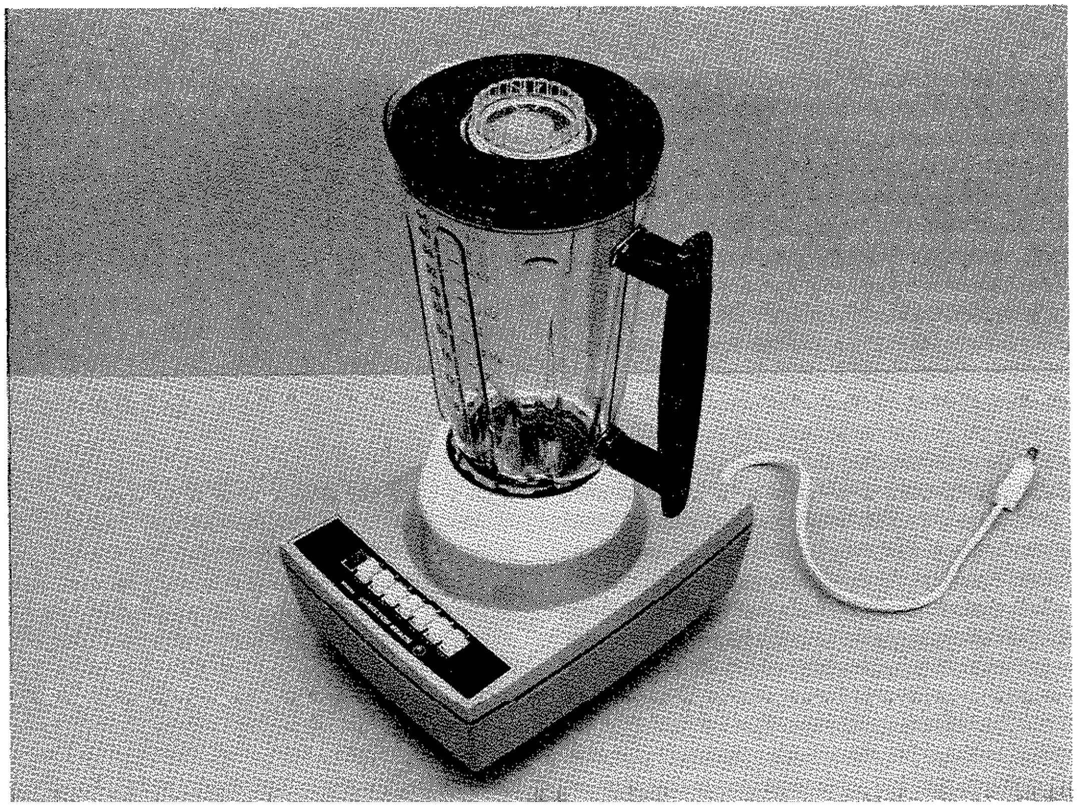


Figure 28. The same object as in Figure 27, but with a viewpoint not parallel to the major components.

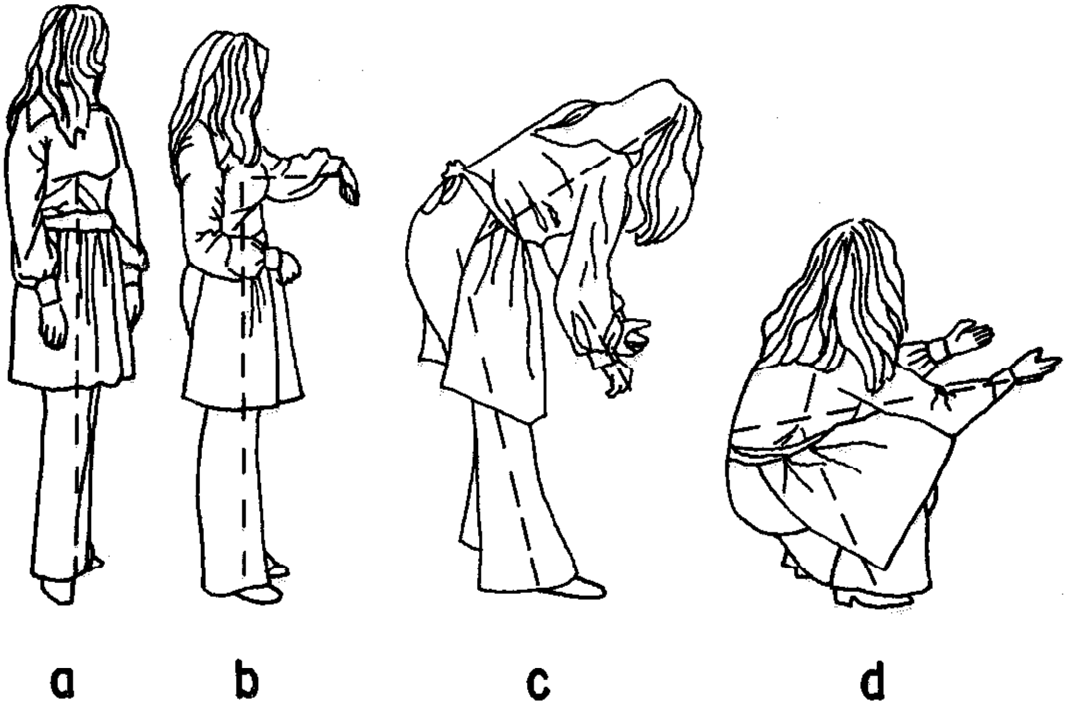


Figure 29. Four configurations of a nonrigid object.

telephone. In that case, it might be either that the relation is "nearby" or else different structural descriptions are necessary for attached and separable configurations. Empirical research needs to be done to determine if less restrictive relations, such as "join" or "nearby," have measurable perceptual consequences. It may be the case that the less restrictive the relation, the more difficult the identifiability of the object. Just as there appear to be canonical views of rigid objects (Palmer et al., 1981), there may be a canonical "configuration" for a nonrigid object. Thus, the poses on the right in Figure 29 might be identified as a woman more slowly than would the poses on the left.

Conclusion

To return to the analogy with speech perception, the characterization of object perception provided by RBC bears a close resemblance to some current views as to how speech is perceived. In both cases, the ease with which we are able to code tens of thousands of words or objects is solved by mapping that input onto a modest number of primitives—55 phonemes or 36 components—and then using a representational system that can code and access free combinations of these primitives. In both cases, the specific set of primitives is derived from dichotomous (or trichotomous) contrasts of a small number (less than ten) of independent characteristics of the input. The ease with which we are able to code so many words or objects may thus derive less from a capacity for coding continuous physical variation than it does from a perceptual system designed to represent the free combination of a modest number of categorized primitives based on simple perceptual contrasts.

In object perception, the primitive components may have their origins in the fundamental principles by which inferences about a three-dimensional world can be made from the edges in a two-dimensional image. These principles constitute a significant portion of the corpus of Gestalt organizational constraints. Given that the primitives are fitting simple parsed parts of an object, the constraints toward regularization characterize not the complete object but the object's components. RBC thus provides, for the first time, an account of the heretofore undecided relation between these principles of perceptual organization and human pattern recognition.

References

Attneave, F. (1982). Pragnanz and soap bubble systems. In J. Beck (Ed.) *Organization and representation in visual perception* (pp. 11–29). Hillsdale, NJ: Erlbaum.

Ballard, D., & Brown, C. M. (1982). *Computer vision.* Englewood Cliffs, NJ: Prentice-Hall.

Barrow, H. G., & Tenenbaum, J. M. (1981). Interpreting line-drawings as three-dimensional surfaces. *Artificial Intelligence, 17,* 75–116.

Bartlett, F. C. (1932). *Remembering: a study in experimental and social psychology.* New York: Cambridge Univ. Press.

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

Bartram, D. (1976). Levels of coding in picture–picture comparison tasks. *Memory & Cognition, 4,* 593–602.

Beck, J., Prazdny, K., & Rosenfeld, A. (1983). A theory of textural segmentation. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 1–38). New York: Academic Press.

Besl, P. J., & Jain, R. C. (1986). Invariant surface characteristics for 3D object recognition in range images. *Computer Vision, Graphics, and Image Processing, 33,* 33–80.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–253). Hillsdale, NJ: Erlbaum.

Biederman, I. (1985). Human image understanding: Recent experiments and a theory. *Computer Vision, Graphics, and Image Processing, 32,* 29–73.

Biederman, I., Beiring, E., Ju, G., & Blickle, T. (1985). *A comparison of the perception of partial vs. degraded objects.* Unpublished manuscript, State University of New York at Buffalo.

Biederman, I., & Blickle, T. (1985). *The perception of objects with deleted contours.* Unpublished manuscript, State University of New York at Buffalo.

Biederman, I., Blickle, T. W., Teitelbaum, R. C., Klatsky, G. J., & Mezzanotte, R. J. (in press). Object identification in multi-object, nonscene displays. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Biederman, I., & Ju, G., (in press). Surface vs. edge-based determinants of visual recognition. *Cognitive Psychology.*

Biederman, I., Ju, G., & Clapper, J. (1985). *The perception of partial objects.* Unpublished manuscript, State University of New York at Buffalo.

Biederman, I., & Lloyd, M. (1985). *Experimental studies of transfer across different object views and exemplars.* Unpublished manuscript, State University of New York at Buffalo.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177.

Binford, T. O. (1971, December). *Visual perception by computer.* Paper presented at the IEEE Systems Science and Cybernetics Conference, Miami, FL.

Binford, T. O. (1981). Inferring surfaces from images. *Artificial Intelligence, 17,* 205–244.

Brady, M. (1983). Criteria for the representations of shape. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 39–84). New York: Academic Press.

Brady, M., & Asada, H. (1984). Smoothed local symmetries and their implementation. *International Journal of Robotics Research, 3,* 3.

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

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Cezanne, P. (1941). Letter to Emile Bernard. In J. Rewald (Ed.), *Paul Cezanne's letters* (translated by M. Kay, pp. 233–234). London: B. Cassirrer. (Original work published 1904).

Chakravarty, I. (1979). A generalized line and junction labeling scheme with applications to scene analysis. *IEEE Transactions PAMI,* April, 202–205.

Checkosky, S. F., & Whitlock, D. (1973). Effects of pattern goodness on recognition time in a memory search task. *Journal of Experimental Psychology, 100,* 341–348.

Connell, J. H. (1985). *Learning shape descriptions: Generating and generalizing models of visual objects.* Unpublished master's thesis, Massachusetts Institute of Technology, Cambridge.

Coss, R. G. (1979). Delayed plasticity of an instinct: Recognition and avoidance of 2 facing eyes by the jewel fish. *Developmental Psychobiology, 12,* 335–345.

Egeth, H., & Pachella, R. (1969). Multidimensional stimulus identification. *Perception & Psychophysics, 5,* 341–346.

Fildes, B. N., & Triggs, T. J. (1985). The effect of changes in curve geometry on magnitude estimates of road-like perspective curvature. *Perception & Psychophysics, 37,* 218–224.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts.* New York: Wiley.

Garner, W. R. (1974). *The processing of information and structure.* New York: Wiley.

Guzman, A. (1968). Decomposition of a visual scene into three-dimensional bodies. *AFIRS Fall Joint Conferences, 33,* 291–304.

Guzman, A. (1971). Analysis of curved line drawings using context and global information. *Machine intelligence* 6 (pp. 325–375). Edinburgh: Edinburgh University Press.

Hildebrandt, K. A. (1982). The role of physical appearance in infant and child development. In H. E. Fitzgerald, E. Lester, & M. Youngman (Eds.), *Theory and research in behavioral pediatrics* (Vol. 1, pp. 181–219). New York: Plenum.

Hildebrandt, K. A., & Fitzgerald, H. E. (1983). The infant's physical attractiveness: Its effect on bonding and attachment. *Infant Mental Health Journal, 4,* 3–12.

Hochberg, J. E. (1978). *Perception* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

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

Humphreys, G. W. (1983). Reference frames and shape perception. *Cognitive Psychology, 15,* 151–196.

Ittleson, W. H. (1952). *The Ames demonstrations in perception.* New York: Hafner.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13,* 289–303.

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Picture and names: Making the connection. *Cognitive Psychology, 16,* 243–275.

Julesz, B. (1981). Textons, the elements of texture perception, and their interaction. *Nature, 290,* 91–97.

Kanade, T. (1981). Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence. 17.* 409–460.

Kanizsa, G. (1979). *Organization in vision: Essays on Gestalt perception.* New York: Praeger.

King, M., Meyer, G. E., Tangney, J., & Biederman, I. (1976). Shape constancy and a perceptual bias towards symmetry. *Perception & Psychophysics, 19,* 129–136.

Lowe, D. (1984). *Perceptual organization and visual recognition.* Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Mark, L. S., & Todd, J. T. (1985). Describing perception information about human growth in terms of geometric invariants. *Perception & Psychophysics, 37,* 249–256.

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

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

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

Marslen-Wilson, W. (1980). *Optimal efficiency in human speech processing.* Unpublished manuscript, Max-Planck-Institut für Psycholinguistik, Nijmegen, The Netherlands.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, Part I: An account of basic findings. *Psychological Review, 88,* 375–407.

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

Neisser, U. (1963). Decision time without reaction time: Experiments in visual scanning. *American Journal of Psychology, 76,* 376–385.

Neisser, U. (1967). *Cognitive Psychology.* New York: Appleton.

Oldfield, R. C. (1966). Things, words, and the brain. *Quarterly Journal of Experimental Psychology, 18,* 340–353.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17,* 273–281.

Palmer, S. E. (1980). What makes triangles point: Local and global effects in configurations of ambiguous triangles. *Cognitive Psychology, 12,* 285–305.

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

Penrose, L. S., & Penrose, R. (1958). Impossible objects: A special type of illusion. *British Journal of Psychology, 49,* 31–33.

Perkins, D. N. (1983). Why the human perceiver is a bad machine. In

J. Beck, B. Hope, & A. Rosenfeld, (Eds.), *Human and machine vision* (pp. 341–364). New York: Academic Press.

Perkins, D. N., & Deregowski, J. (1982). A cross-cultural comparison of the use of a Gestalt perceptual strategy. *Perception, 11,* 279–286.

Pomerantz, J. R. (1978). Pattern and speed of encoding. *Memory & Cognition, 5,* 235–241.

Rock, I. (1983). *The logic of perception.* Cambridge, MA: MIT Press.

Rock, I. (1984). *Perception.* New York: W. H. Freeman.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Ryan, T., & Schwartz, C. (1956). Speed of perception as a function of mode of representation. *American Journal of Psychology, 69,* 60–69.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three dimensional objects. *Science, 171,* 701–703.

Sugihara, K. (1982). Classification of impossible objects. *Perception, 11,* 65–74.

Sugihara, K. (1984). An algebraic approach to shape-from-image problems. *Artificial Intelligence, 23,* 59–95.

Treisman, A. (1982). Perceptual grouping and attention in visual search for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 194–214.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Trivers, R. (1985). *Social evolution.* Menlo Park, CA: Benjamin/Cummings.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113,* 169–193.

Ullman, S. (1984). Visual routines. *Cognition, 18,* 97–159.

Virsu, V. (1971a). Tendencies to eye movement, and misperception of curvature, direction, and length. *Perception & Psychophysics, 9,* 65–72.

Virsu, V. (1971b). Underestimation of curvature and task dependence in visual perception of form. *Perception & Psychophysics, 9,* 339–342.

Waltz, D. (1975). Generating semantic descriptions from drawings of scenes with shadows. In P. Winston (Ed.), *The psychology of computer vision* (pp. 19–91). New York: McGraw-Hill.

Winston, P. A. (1975). Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 157–209). New York: McGraw-Hill.

Witkin, A. P., & Tenenbaum, J. M. (1983). On the role of structure in vision. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 481–543). New York: Academic Press.

Woodworth, R. S. (1938). *Experimental psychology.* New York: Holt.

Received June 28, 1985
Revision received June 3, 1986 ■