Journal of Experimental Psychology:
Human Perception and Performance
1993, Vol. 19, No. 6, 1162-1182

0096-1523/93/$3.00

Recognizing Depth-Rotated Objects: Evidence and Conditions for Three-Dimensional Viewpoint Invariance

Irving Biederman and Peter C. Gerhardstein

Five experiments on the effects of changes of depth orientation on (a) priming the naming of briefly flashed familiar objects, (b) matching individual sample volumes (geons), and (c) classifying unfamiliar objects (that could readily be decomposed into an arrangement of distinctive geons) all revealed immediate (i.e., not requiring practice) depth invariance. The results can be understood in terms of 3 conditions derived from a model of object recognition (I. Biederman, 1987; J. E. Hummel & I. Biederman, 1992) that have to be satisfied for immediate depth invariance: (a) that the stimuli be capable of activating viewpoint-invariant (e.g., geon) structural descriptions (GSDs), (b) that the GSDs be distinctive (different) for each stimulus, and (c) that the same GSD be activated in original and tested views. The stimuli used in several recent experiments documenting extraordinary viewpoint dependence violated these conditions.

Consider Figure 1. The viewer readily appreciates that it shows two different views of the same object, despite myriad differences in the two silhouettes and in the local image features (namely, vertices, lines, and length and curvature of these lines). In general, people typically evidence little difficulty in recognizing a familiar object when they view that object from a different perspective in depth.

Is Depth Invariance Achieved Through Familiarity With Specific Views or Invariant Primitives?

Recent theoretical proposals about how such viewpoint invariance might be achieved fall into two classes. One view (e.g., Edelman & Bülthoff, 1992; Poggio & Edelman, 1990; Rock, 1973; Tarr, 1989) contends that the invariance is based on familiarity: Separate visual representations, typically templates, are created for each experienced viewpoint. If an image does not match a previously experienced image for which there exists a representation, then classification is accomplished by a process (e.g., mental rotation, interpolation, or extrapolation) that incurs a cost in time proportional to the angular difference between the image and the closest studied view. According to these theories, the reason the chair seen on the left in Figure 1a is identified as the same object as the chair seen in Figure 1b is that either both views are stored in memory, with a link to a common higher level representation of the chair, or that one view is stored and the other is rotated (or extrapolated) to match the stored view.

Three major empirical observations can be cited that provide apparent support for this theoretical position of view-specific templates: (a) with familiar objects, a cost in recognition speed is seen when an object is shown at a new viewpoint (Bartram, 1974); (b) subjects display abysmal performance when they attempt to recognize certain kinds of unfamiliar objects depicted from a new orientation (e.g., Edelman, Bülthoff, & Weinshall, 1993; Rock & DiVita, 1987; Tarr, 1989); and (c) the effect on performance of stimulus rotation angle from the originally studied orientation declines as intervening views are experienced and stored in memory (Edelman et al., 1993; Tarr, 1989).

There is a problem, however, with the view-specific account of recognition of familiar objects: Although the class of objects represented by an object might be familiar, the exact contours present in an image of a familiar object, such as those in Figure 1, are unfamiliar. It is unlikely that the reader has previously experienced the particular arrangement of contours shown in either of these images. Thus, it is not obvious how familiarity could have endowed these particular images with viewpoint invariance.

An alternative theoretical proposal to one based on familiarity with specific images is that depth invariance, up to the accretion and deletion of parts, can be achieved in the representation derived from a single view of the object. One way that this might be done is to represent the image as a structural description specifying the relations among viewpoint-invariant volumetric primitives (e.g., geons), as assumed by geon theory (Biederman, 1987; Hummel & Biederman, 1992). As long as two views of an object activate the same structural description, which is possible because the same geons can be activated by different local image features, viewpoint invariance is expected.

Irving Biederman, Department of Psychology, University of Southern California, Los Angeles; Peter C. Gerhardstein, Department of Psychology, University of Minnesota, Twin Cities.

This research was supported by grants from the U.S. Air Force Office of Scientific Research (Grants 88-0231 and 90-0274) and the McDonnell-Pew Cognitive Neuroscience Program (Grant T89-01245-029). This article benefited greatly from the reviews and comments of Eric E. Cooper, Stephen Palmer, James Pomerantz, and Dan Kersten.

Correspondence concerning this article should be addressed to either Irving Biederman, Department of Psychology, Hedco Neuroscience Building, University of Southern California, Los Angeles, California 90089-2520 (electronic mail: ib@rana.usc.edu), or Peter C. Gerhardstein, who is now at the Department of Psychology, University of Arizona, Tucson, Arizona 85721 (electronic mail: gerhard@ccit.arizona.edu).

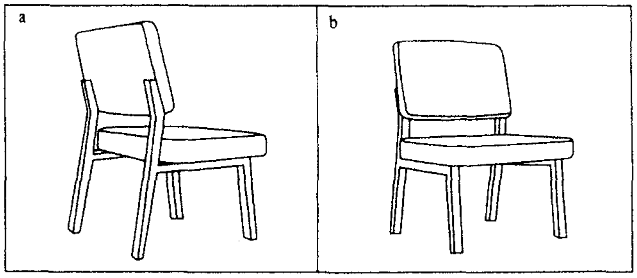

Figure 1. Two views of a chair. (The pose depicted in Panel a is a 90° clockwise rotation of the pose depicted in Panel b.)

Hummel and Biederman (1992) also offered an extensive analysis and model of why, paradoxically, sizable costs in object recognition speed and accuracy have been observed for planar rotation (Jolicoeur, 1985). Such an effect is termed paradoxical because image characteristics such as spatial frequency and lines and vertices are unaffected by planar rotation but are often drastically altered by rotation in depth. Briefly, given a structural description specifying geons, their individual attributes, and the relations between the geons, rotation in the plane alters the TOP-OF, BOTTOM-OF, and SIDE-OF values of the relations between geons and the VERTICAL, HORIZONTAL, and OBLIQUE values for the attributes of the individual geons. LEFT-OF and RIGHT-OF, which can be affected by sizable rotation in depth, are assumed not to be represented in the structural description for object recognition but are instead specified by the viewer-centered representations of the dorsal visual system presumed to mediate motor interaction (Biederman & Cooper, 1992). Consistent with the assumption that LEFT-OF and RIGHT-OF are not specified for object recognition is the lack of any effect of reflection on priming in object naming tasks (Biederman & Cooper, 1991a, 1991b). Most of the objects in the Biederman and Cooper experiments were bilaterally symmetrical, so their reflection would have been equivalent to a 180° rotation in depth.[1]

Problems in Obtaining Invariance

Invariance over rotation in depth would not be the only invariance expected to be characteristic of a system designed to achieve entry-level object recognition.[2] Biederman and Cooper (1992) argued that in addition to invariance in depth, it would also be highly desirable for representations of shape for recognition to be invariant with respect to retinal position and size in that it would be rare for a subsequent encounter with a previously experienced object to duplicate the original position, size, and orientation in depth of the object. Biederman and Cooper (1992) speculated that these invariances are what distinguish the ventral cortical visual system (Ungerleider & Mishkin, 1982), presumed to accomplish recognition, from the dorsal cortical visual system, which they argued supports representations subserving motor interaction. Competent motor interaction requires specification of the position, size, and orientation in depth of the object. Biederman and Cooper (1991a, 1992; see also Cooper, Biederman, & Hummel, 1992) documented strong invariance over changes in size and position in the priming of entry level object naming. The invariance was termed "strong" in that there was no reduction in the magnitude of priming when the size or position of a picture was changed from its initial presentation. A central empirical goal of the present effort was to assess the conditions under which strong invariance might be evidenced over rotation in depth.

Although the activation of a representation of an object for recognition may be invariant over orientation disparity, it is possible that performance itself does not reveal the invariance. Actually obtaining invariance in overt responding would require that the task reflect only the latencies of activation of ventral object representations. Biederman and Cooper (1992) and Cooper et al. (1992) reported that the changes in image size or position that had no effect on the priming of picture-naming performance produced considerable interference with old–new recognition memory judgments. They argued that the latter task was controlled by feelings of familiarity that were influenced by both the dorsal and ventral systems. Naming, however, was only influenced by ventral system representations, which are presumably invariant over changes in size, position, and orientation. Humphrey and Kahn (1992) recently reported that depth-orientation changes did interfere with old–new recognition memory of unfamiliar objects. Similarly, the frequently studied mental rotation task in which mirror reflections of a rotated shape have to be distinguished (e.g., Shepard & Cooper, 1983) likely reflect dorsal system functioning (Kosslyn, in press; Kosslyn & Koenig, 1992). The experiments reported in the present investigation involved paradigms designed to reduce the reliance on feelings of familiarity or the need to distinguish between mirror reflections of the same object. The possibility that responses in a given task might not be solely influenced by invariant representations leads to an asymmetry in what can be concluded from experiments assessing invariance: The absence of an effect of orientation disparity indicates that the representation was invariant (assuming sufficient statistical power), but an effect of orientation disparity could be the result of an influence from another system.

[1] The Hummel and Biederman (1992) simulation also produced a heretofore unexplained effect in the Jolicoeur (1985) experiment: The deleterious effect of planar rotation on naming reaction times (RTs), which monotonically increased from 0° to 120°, were reversed as the rotation angle increased from 120° to 180°, so that RTs were shorter at 180° than at 120°. This occurred in the Hummel and Biederman (1992) simulation because at 180° the orientation of the individual geons was restored so that a geon that was VERTICAL (or HORIZONTAL or OBLIQUE) at 0° was again VERTICAL (or HORIZONTAL or OBLIQUE) at 180°.

[2] In discussing the role of shape in object classification, "entry level" is preferred over "basic level" in that it allows members of a class that have atypical shapes (e.g., penguins and ostriches for the class, birds) to have a classification level equivalent to the standard basic level (Jolicoeur, Gluck, & Kosslyn, 1984). That is, penguins and ostriches are likely classified as penguins or ostriches before they are classified as birds.

When Should Viewpoint Invariance From a Single Pose of an Object Be Expected?

From the perspective of geon theory, it is the structural description (consisting of geons, their attributes, and their relations with adjacent geons) that allows the viewpoint invariance: If two views of an object activate the same structural description, then they should be treated as equivalent by the object recognition system (Biederman, 1987; Hummel & Biederman, 1992).

Not all objects will activate a structural description that remains invariant over a large change of orientation and not all sets of objects have distinctive structural descriptions for their individual members. Moreover, for viewpoint invariance to be achieved, different views of an object would have to activate the same structural description. According to the Hummel and Biederman (1992) neural net implementation of geon theory, strong, immediate viewpoint invariance would not be expected unless all three conditions are met.[3] These conditions also specify the conditions under which entry-level classes will be defined on the basis of object shape.[4] The conditions will be considered in turn.

Condition 1: Geon Structural Descriptions (Readily Identifiable Invariant Parts)

Immediate viewpoint invariance requires that an object must be decomposable into viewpoint invariant parts (e.g., geons) so that a geon structural description (GSD) that specifies the geons and their relations can be activated. If the contours of the object cannot be readily decomposed into geons, then the representation will not be one that allows viewpoint-invariant recognition. Rock and DiVita's (1987) lumpy clay mass and crumpled paper are examples of objects that do not meet this condition. Neither type of object can be recognized when rotated in depth. Generally, a GSD will fail to be activated because the object does not readily decompose into parts or the parts are highly irregular, corresponding more to texture regions (if they are numerous) than volumetric entities.

Condition 2: Distinctive Geon Structural Descriptions for Different Stimuli

Even if each member of a set of stimuli readily activates a GSD, viewpoint invariance will not be achieved unless each stimuli has a different GSD. That is, the GSDs for different objects must be distinctive. A measure of GSD similarity is proposed later in this section. Tarr's (1989) stimuli, two of which are depicted in Figure 2, are an example in which Condition 2 is not met. Each of Tarr's objects were made up of seven bricks, varying in length, with each brick connected in orthogonal, END-TO-END or END-TO-MIDDLE relations to other blocks.[5] The GSDs for the different objects in Tarr's stimulus set, from the perspective of the Hummel and Biederman (1992) model, would have been virtually identical.

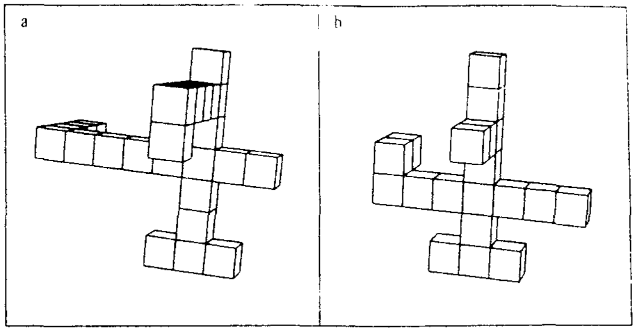

Figure 2. Redrawings of two stimuli from Figure 1 of Tarr (1989). (Note that although the two are not of the same object, they have almost all the same parts. A description such as the following would be necessary to distinguish them from each other and from the other five members of the stimulus set: a midlength horizontal brick, end-connected to the long brick and above another long brick and perpendicular to a short horizontal brick. From *Orientation Dependence in Three-Dimensional Object Recognition*, p. 23, Figure 1, by M. J. Tarr, 1989, unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA. Copyright 1989 by M. J. Tarr. Adapted by permission.)

Condition 3: Identical Geon Structural Descriptions Over Different Viewpoints

If two views of the same object activate the same GSD, and the prior conditions are met, then strong viewpoint invariance will be obtained, subject to the condition that the task not allow other (viewpoint-dependent) systems to affect performance, as discussed previously. It is possible that different geons will be present in two views of an object, because of accretion (revelation) and deletion (occlusion), as when looking at the front versus the back of a house. Different GSDs would then be activated and viewpoint invariance would not be expected. In a sense, this condition is identical to the previous (distinctiveness) condition, in that if two views of an object have different GSDs, then viewpoint invariance would not be expected.

Objects differ in the degree to which they retain the same part structure over rotation. It is also possible that a part

[3] The invariance is termed *immediate* if it can be obtained without practice or familiarity with the specific stimuli.

[4] The scope of the conditions is not limited to entry-level classifications but rather includes all shape classifications that can be performed easily. These are likely to include the majority of the subordinate classifications made in our daily lives (cf. Biederman & Shiffrar, 1987).

[5] Although JOIN–TYPE (END-to-END, END-to-MIDDLE) was specified in Biederman's (1987) original proposal, it was not explicitly included in the Hummel and Biederman (1992) implementation (which was not presented as an exhaustive account of object recognition). However, such a relation could readily be incorporated explicitly into the existing architecture with several additional units. The model will, however, often distinguish between END-to-END and END-to-MIDDLE joins through differential activation of the units for relative position of the geons as OBLIQUE, ORTHOGONAL, or PARALLEL.

structure of some kind is activated (perhaps with weak activation of geons) by an image of an object "accidentally," in the sense that those parts can be discerned from only a small range of viewing orientations. The criterion for an accident is that small changes in orientation result in large changes in the part structure, despite the presence of the same contours in the different views. Wire frame constructions readily lead to such accidents as occurred, for example, with the curved wire frame constructions of Rock and DiVita (1987), depicted in Figure 3, and the bent "paper clips" of Edelman, Bülthoff, and Weinshall (1989), depicted in Figure 4. Figures 3a and 3c are 50° rotations in depth (in different directions) from Figure 3b of Rock and DiVita's (1987) object. Approximately the same part structure can be discerned in Figures 3a and 3b. However, that structure is not apparent in Figure 3c, so this image would fail to meet Condition 3 with respect to the images in Figure 3a or 3b. Both the Rock and DiVita (1987) and Edelman et al. (1993) stimuli are considered in detail in the critical review section of the Introduction.

Condition 3 can be subsumed under the general topic of aspect graphs (Koenderink, 1990). An aspect graph specifies the features that are present in an image from a given viewpoint and maps how the feature set changes as viewpoint changes. The set of possible views can be conceptualized as points on the surface of a transparent viewing sphere containing the object. As the viewpoint changes for any object (but a sphere), qualitative changes in image features occur. For example, the rotation of a brick changes an arrow vertex, when two surfaces are in view, to an L vertex, when one of the surfaces is occluded. Every such qualitative change produces a new aspect, each of which can be conceptualized as a patch (set of points) on the surface of the viewing sphere containing the loci of viewpoints from which the same features of the object can be seen. Most of the work on aspect graphs has defined features in terms of local contours, such as vertices, lines, and inflection points (e.g., Eggert & Bowyer, 1990; Kriegman & Ponce, 1990). The problem with such a local definition of a feature is that when all possible viewpoints of a complex object are determined, then hundreds, if not thousands, of different aspects can result. Local features thus define a representation that is relatively unstable over viewpoint. If the features are geons, however, then far fewer aspects are required, in that many different sets of features can map onto the same geon (Biederman, 1987; Dickinson, Pentland, & Rosenfeld, 1992; Hummel & Biederman, 1992).

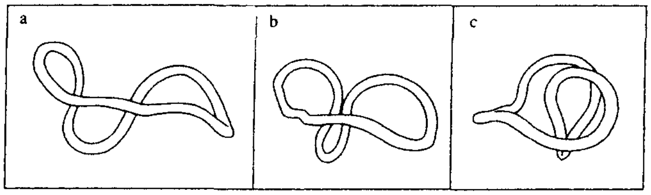

Figure 3. A line drawing of a wire-frame object similar to one shown in Figure 1 of Rock and DiVita (1987). (The three images are of the same object, differing by a 50° rotation in depth. From "A Case of Viewer-Centered Perception," by I. Rock and J. DiVita, 1987, *Cognitive Psychology, 19,* p. 282. Copyright 1987 by I. Rock. Adapted by permission.)

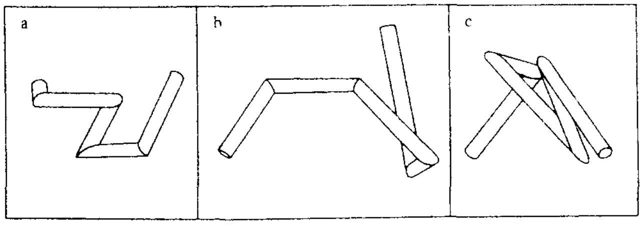

Figure 4. Line drawings of three poses of an object like those in Figure 1 of Edelman, Bülthoff, and Weinshall (1989). (The poses differ by about 50°. This object had to be distinguished from nine others, all made up of the same set of five wires differing only in angle. From *Stimulus Familiarity Determines Recognition Strategy for Novel 3D Objects* [Artificial Intelligence Laboratory Tech. Rep. No. 1138, p. 2], by S. Edelman, H. Bülthoff, and D. Weinshall, 1989, Cambridge, MA: MIT. Copyright 1989 by S. Edelman. Adapted by permission.)

Criterion for Distinctiveness

A possible measure of distinctiveness for Condition 2 is suggested by the neural net implementation of recognition-by-components by Hummel and Biederman (1992). The input to the sixth layer of the model is a pattern of activation across 21 units specifying one geon, its attributes, and its relations. Specifically, the units specifying a distributed representation for a given geon type (e.g., brick or cylinder) of an object are bound (by simultaneous firing) to units representing: (a) two attributes of that geon: orientation (VERTICAL, HORIZONTAL, OBLIQUE) and coarsely coded aspect ratio; and (b) three relations between adjacent geons: vertical position (TOP-OF, BELOW, BESIDES), size (LARGER, SMALLER, SAME) and relative orientation (ORTHOGONAL, OBLIQUE, PARALLEL). Patterns of activity over the 21 units in successive time slices (or subslices) represent the different geons (and, for each, their attributes and relations) of the object. Each of these patterns recruits a separate unit, termed a *geon feature assembly* (GFA). Units in the last (seventh) layer are recruited from the integration of the output of successive GFAs from layer six to form object representations.

The measure of object similarity can be derived from the 21-cell vectors representing each GFA. For each pair of objects, the sum of the absolute differences of corresponding cells, the Hamming distance (Ullman, 1989), provides a measure of similarity. As a simplified example, consider two 5-cell vectors with values of 1,0,0,1,1 and 0,0,1,1,0. Their Hamming distance would be three (as they have different values in cells 1, 3, and 5).[6] For the measure of similarity, the GFA vectors should be ordered so as to minimize the Hamming distance. If the different objects comprising a set

[6] In this example, the cells are assumed to have equal weights. A more appropriate measure would be to multiply them by the connection weights in the model.

of stimuli differ in their geons, attributes, and relations, then there will be a large Hamming distance between the objects. With such stimuli, the Hummel and Biederman (1992) network attained viewpoint invariance over rotation in depth. When the Hamming distance is small between the members of a set of stimuli, invariance is less likely.[7]

There are at least two reasons why strong depth invariance may be lost with highly similar stimuli such as those of Tarr (1989). If the only distinguishing information is viewpoint dependent, such as the length (or aspect ratio) of a part, the two-dimensional projection of that length (or aspect ratio) will vary as the object is rotated in depth, requiring additional processing to recover the original metric values. Second, with highly similar stimuli, the subject may use salient left–right, viewer-centered information, for example, information that a given part was on the right side. Such a strategy would lose invariance over mirror-image reflection and may be more characteristic of dorsal than ventral visual system processing, as discussed earlier.

In the context of the previously specified conditions for viewpoint invariance, the next section presents a critical review of studies that failed to obtain invariance.

A Critical Review of the Evidence for Template Familiarity as a Basis for Rotational Invariance

Findings of enormous cognitive costs when viewing objects at novel orientations (Edelman & Bülthoff, 1992; Rock & DiVita, 1987; Tarr, 1989) appear, at first blush, to challenge the central motivation of theories that posit viewpoint-invariant representations, such as geon theory. However, in attempting to create unfamiliar stimuli, these investigators produced sets of objects that failed to meet at least one of the conditions for viewpoint invariance, either because the stimuli did not decompose into a GSD or because the set members did not activate distinctive GSDs or produced nonstable part structures. Thus, the members of the stimulus sets used in each of these experiments would be unlikely to fall into different entry-level classes in any culture, according to geon theory. The reason for this requires a closer look at the experiments.

Nondecomposable, Nondistinctive Stimuli

The first and perhaps the most striking demonstration of viewpoint dependence was that of Rock and DiVita (1987). Their subjects studied a series of smoothly curved wire-frame objects for 4 s at a given orientation, such as that shown in Figure 3b, which shows a line drawing rendition of one of their stimuli. They then tested recognition when the objects were viewed at diagonal rotations in depth, as illustrated in Figure 3c. (Subjects actually viewed each object in one of the quadrants of the visual field, for example, the upper left. They were then tested for recognition of the object when it was presented in another quadrant.) The remarkable result was that Rock's subjects could recognize the original object only 39% of the time against similar smoothly curved (real) wire-frame objects, when viewed at the new orientation. (Objects in their original positions could be recognized 75% of the time.) Indeed, Rock, Wheeler, and Tudor (1989) showed that people could not even imagine how smoothly curved wire-frame objects look when rotated. The inability to recognize objects from a different viewpoint was not limited to wire-frame objects. Rock and DiVita (1987) presented a demonstration of two views of a blob-like clay construction and crumpled paper in which it is subjectively clear that it would be extremely difficult to determine that the same object was projecting each image.

For all of Rock and DiVita's (1987) objects, the relative depth of each point on the object could be accurately determined (Rock et al., 1989), so it was not the case that the difficulty was a consequence of an input that was initially indeterminate with respect to its three-dimensional structure. Rock and DiVita's demonstrations are important in that they show that even with accurately perceived depth, a viewpoint-invariant representation may not be possible for an object.

Rock and DiVita's (1987) wire-frame objects (Figure 3) failed all three conditions for invariance. The smoothly curved continuous segments led to only a weak part structure: The concavities at the matched T-vertices, which suggest these "parts," are not in the object but are accidents of viewpoint in the image. These concavities define loops that would likely correspond to the object's most salient parts in the two-dimensional image. The stimuli also failed Condition 3 (identical GSDs over rotation) in that the loops could change over modest rotations in depth, as noted previously with Figures 3b and 3c, because slight changes in viewpoint could create or eliminate the concavities, causing the part structure to be altered even though the contour that comprised the loop remained in view. The curved segments were highly similar over the members of the set of stimuli (all smoothly curved wire-frame objects) thus not satisfying Condition 2 either.

Whereas the wire-frame objects might have had accidental parts, Rock and DiVita's (1987) clay blobs and crumpled newspapers and Bülthoff and Edelman's (1992) blob-like, ameboid volumes did not even have structures that could be readily decomposed into parts. Differences among such stimuli are defined in terms of viewpoint-dependent metric values for length (or aspect ratio) and degree of curvature. The wadded paper actually presented an additional factor: People tend to interpret a large number of repeated elements in any display as a texture field in which the relations among the individual elements are not specified as part of the representation (Beck, 1967; Hummel & Biederman, 1992; Treisman, 1988). Not surprisingly, immediate viewpoint invariance with such stimuli was not obtained.

[7] Somewhat consistent with this argument is a recent article by Edelman (1993), in which he reported viewpoint invariance for the classification of dissimilar images into two categories (dog vs. monkey), but not for the classification of similar images. Although Edelman generated the various values of his stimuli metrically, his dissimilar stimuli differed in nonaccidental contrasts (e.g., pointed ears vs. round ears) and thus would be expected to yield viewpoint invariance from the perspective of recognition-by-components.

Decomposable but Nondistinctive Stimuli

Tarr (1989) taught subjects names for three of a set of seven objects constructed entirely from five or six bricks made up of varying numbers of small cubes (as illustrated in Figure 2) and then had the subjects identify experimental images as one of the three named objects or as an unnamed member of the set when the objects were shown at various orientations. It is important to note that an extensive training regimen was required for the subjects to master the set of stimuli. Among other techniques for training, the subjects duplicated each object five times from a set of toy blocks. Relying on their memory, they then had to build the objects twice correctly before they could start the experiment. When naming the objects at new orientations, there was initially a large effect on reaction times (RTs) as a function of changes in depth orientation (from the original, studied orientation), which produced a relatively large effect of angular disparity of 167°/s.[8] This effect decreased to 588°/s (i.e., rotation speed increased) after 12 blocks of trials, each with 24 exposures of each object at various concentrations (288 trials total). Tarr then tested the subjects with a previously unseen "surprise" orientation. This manipulation resulted in a reduction of speed to 119°/s for the surprise block, indicating that the subjects had not learned a general skill for rapid mental rotation.

Tarr (1989) argued that his subjects were using a multiple-views plus rotation strategy in which the subjects improved over the first 12 blocks of trials as they became more familiar with the views of the objects that they saw; however, this improvement was specific to the views with which the subjects were familiar. Tarr believed that the results of the surprise-view test demonstrated that the representation was specific to familiar views, because if the subjects had developed an invariant representation, then seeing the surprise view would not have resulted in the increased effect of orientation disparity that he observed.

As noted previously, Tarr's (1989) stimuli meet Condition 1 (readily identifiable parts) but they fail Condition 2 (GSD distinctiveness) in that all of his objects comprised the same geons (bricks), having approximately the same variations in length, with the same relations between the parts (ORTHOGONAL, END-TO-MIDDLE, or END-TO-END connections). Consequently, the Hamming distances between GSDs of different objects in Tarr's set would have been small, according to the Hummel and Biederman (1992) model. Tarr's objects were distinguishable only by way of a highly complex descriptor specifying the relations between parts in terms of the lengths of two or more other parts in the object. The objects can be parsed into viewpoint-invariant volumes, but the resulting descriptions did not differ (across objects) by the geons, geon attributes, and relations that geon theory specifies are essential for invariant recognition. Thus, unlike recognition of familiar objects from different entry-level categories, distinguishing among members of this stimulus set would be difficult, and long recognition times and large effects of depth orientation change would be expected.

Edelman et al. (1989) conducted a set of experiments with 10 objects, each constructed from 5 wirelike cylinders connected end-to-end, with the connection angle perturbed by varying amounts, which resembled bent paper clips extending in depth, one of which is depicted in Figure 4. (The high similarity of a set of 10 such objects, with the formidable recognition problem that they pose to a human subject, can be appreciated by considering how difficult the objects in Figure 14 [presented later] would be to distinguish if they all had another thin cylinder as their central geon.) Their subjects performed a task similar to Tarr's (1989), in which they had to determine whether a given stimulus was a target or a distractor. This task was virtually impossible to perform initially, but after extensive practice (14,400 trials per subject) a speed of approximately 500°/s was achieved. This level of performance was similar to that attained by Tarr's subjects after 12 blocks of practice. Edelman et al.'s (1989) stimuli, like Tarr's, satisfied Condition 1, in that the objects could be decomposed into viewpoint invariant parts, but failed Condition 2, in that the stimuli had highly similar structural descriptions. In addition, the Edelman et al. stimuli also failed Condition 3 in that slight variations in viewpoint led to dramatically different relations among the wires.

It is tempting to consider the processing revealed in the Tarr (1989) and Edelman et al. (1989) experiments as characteristic of subordinate-level classification (e.g., how we might know that an object is a particular model of a car) rather than entry-level processing (e.g., that it is a car). Although we defer to the Discussion section an account of the relevance of such processing for typical human behaviors, we note here that in distinguishing among the myriad objects and creatures in our daily lives, we are almost never required to undertake the extraordinarily difficult metric processing suggested by Bülthoff and Edelman (1992).

Familiar Objects

Rock and DiVita (1987) used unfamiliar stimuli, as was the case in the other experiments that showed extreme difficulty in perceiving depth-rotated images. According to Edelman and Bülthoff (1992), Edelman and Weinshall (1991), and Tarr (1989), this characteristic (familiarity) is critical for the apparent ease demonstrated in everyday recognition of objects presented at novel orientations.

An important question bearing on this problem is how well familiar objects are, in fact, recognized over changes in depth orientation. The one published study of this issue, Bartram (1974), found only partial evidence for viewpoint invariance.

[8] The standard format in research on the effects of orientation disparity is to present the effects of rotation angle on RTs in terms of milliseconds per degree. Because this value would be so small in most of the results considered here (fractions of a millisecond per degree), we have opted to describe costs of angular disparity in terms of speed, as degrees per second. Higher values thus indicate smaller effects of rotation (i.e., faster speeds). In using the term rotation speed, no theoretical interpretation (viz., that rotation has a psychological or physical analog) is intended.

In Bartram's priming experiments, subjects named black-and-white photographs of familiar objects from different entry-level classes. A set of images was presented in an initial block of trials, followed by images in one of four conditions in seven subsequent blocks. The conditions were: (a) identical pictures; (b) rotated pictures (the same objects photographed from eight spatial viewpoints approximately 45° apart from each other); (c) different exemplars with the same name (e.g., two different kinds of chairs); and (d) different objects with different names. The dependent variable was the magnitude of reduction in RT over the eight trial blocks.

Performance was best with the identical pictures, which were named faster than the different exemplar pictures, which in turn were named faster than the different objects pictures. The critical question for the issue at hand was the performance with the rotated pictures. If their recognition was viewpoint invariant, they should have been named as quickly as the identical pictures. However, RTs to the rotated pictures were not only slower than to pictures in the identical condition but were almost as slow as those in the different exemplar condition.

Several features of Bartram's (1974) design require a reexamination of the recognition of familiar objects across changes in depth orientation. First, it is quite likely that accretion and deletion of parts occurred between the different orientations in the rotated condition. Second, it would have been impossible to have every picture appear in every condition. Thus, it is possible that some of the results might have been a function of the various orientations differing in "canonicality" (Palmer, Rosch, & Chase, 1981).

Overview of Experiments

Experiments 1 and 2 were designed to examine viewpoint invariance with stimuli that were less likely than Bartram's (1974) to change parts across views, while balancing presentation orders to control for possible effects of canonical views. In contrast to Bartram's results, these experiments document strong invariance over rotation in depth—as long as the same, distinctive GSD can be activated in the two views (i.e., the three conditions for invariance are satisfied). Experiment 3 used unfamiliar (nonsense) objects in a sequential same–different object matching task. Unlike previous sets of unfamiliar stimuli, these objects did meet Conditions 1 and 2 for obtaining invariance. On a *same* trial, the second object could be depicted at different orientations in depth from its pose of the first trial, sometimes with the same geons in view (thus satisfying Condition 3) and sometimes with occlusion and accretion of a geon (thus failing to satisfy Condition 3). These stimuli showed virtually no effect of rotation in depth as long as the same geons were present in the two images. An implicit assumption underlying the conditions is that the geons, individually, are immediately viewpoint invariant. Experiment 4 documented that this is the case in that the detection of single geons showed no effect of orientation disparity. Experiment 5 showed that the addition of a single distinctive geon to the otherwise highly viewpoint-dependent Edelman et al. (1989) paper clip objects, conferred immediate depth invariance to the recognition of those objects.

Experiment 1: Priming Familiar Objects

Experiment 1 used a priming paradigm that assessed the effects of naming a briefly presented picture on the speed and accuracy of naming subsequent related pictures that were also briefly presented. After exposure to a set of objects at a particular orientation, the set was shown again, with some objects at the same orientation and some at differing orientations in depth. An effort was made to select poses at the various orientations for a given object so that the same parts would be in view. Half of the images at each second-block orientation were of the same physical object, whereas half of them were of different objects with the same entry-level name, similar to Bartram's (1974) different-exemplar condition. Figure 5 illustrates the conditions.

An advantage of the same over the different-exemplar condition is that it documents that some of the priming was perceptual and not just due to activation of the basic level name or concept.[9] With respect to the rotation variable, the absence of an effect of a change in orientation would be evidence for viewpoint-invariant recognition.

Method

Subjects. The subjects were 48 native English-speaking individuals with normal or corrected-to-normal vision who participated for payment or credit in Introductory Psychology at the University of Minnesota. All were able to achieve a preset criterion (established from the error percentages present in previous studies using similar stimuli) of not exceeding 10% errors over the course of the experiment.

Stimuli. Forty-eight line drawings—two exemplars each of 24 different objects having a common, readily available basic name (e.g., a passenger jet airplane and a propeller airplane, both of which would be called an "airplane" or "plane")—were drawn on a Macintosh II through a three-dimensional drawing package (Swivel 3D; Paracomp, San Francisco, California). Three views of each object, differing by 67.5°, corresponding to three views about the vertical axis in a right-handed world coordinate system, were exported from this package. The orientation of some images was altered by up to ±5° to reduce the incidence of part changes or extensive foreshortening of a part from one orientation to the next. Even with this flexibility, it proved to be impossible to eliminate all such part changes and foreshortening.) The images were then redrawn with a line width of 2 pixels and saved in the standard Macintosh PICT format with Adobe Illustrator 88 (Adobe Systems, Mountain View, California).

The view of each object that presented the largest span (in any single two-dimensional direction) was scaled to fit a circle the diameter of which subtended a visual angle of 4.5°. The other views of each object were resized quantitatively by the same amount as the first view so as to preserve their size relative to the first view.

[9] It should be noted that the advantage of the same exemplars over different ones provides only a lower bounds estimate of the visual (vs. nonvisual) component of priming in that the different exemplars for a basic level object class are almost always more similar to each other in shape than they are to arbitrarily selected images from different basic-level categories (Biederman & Cooper, 1991b).

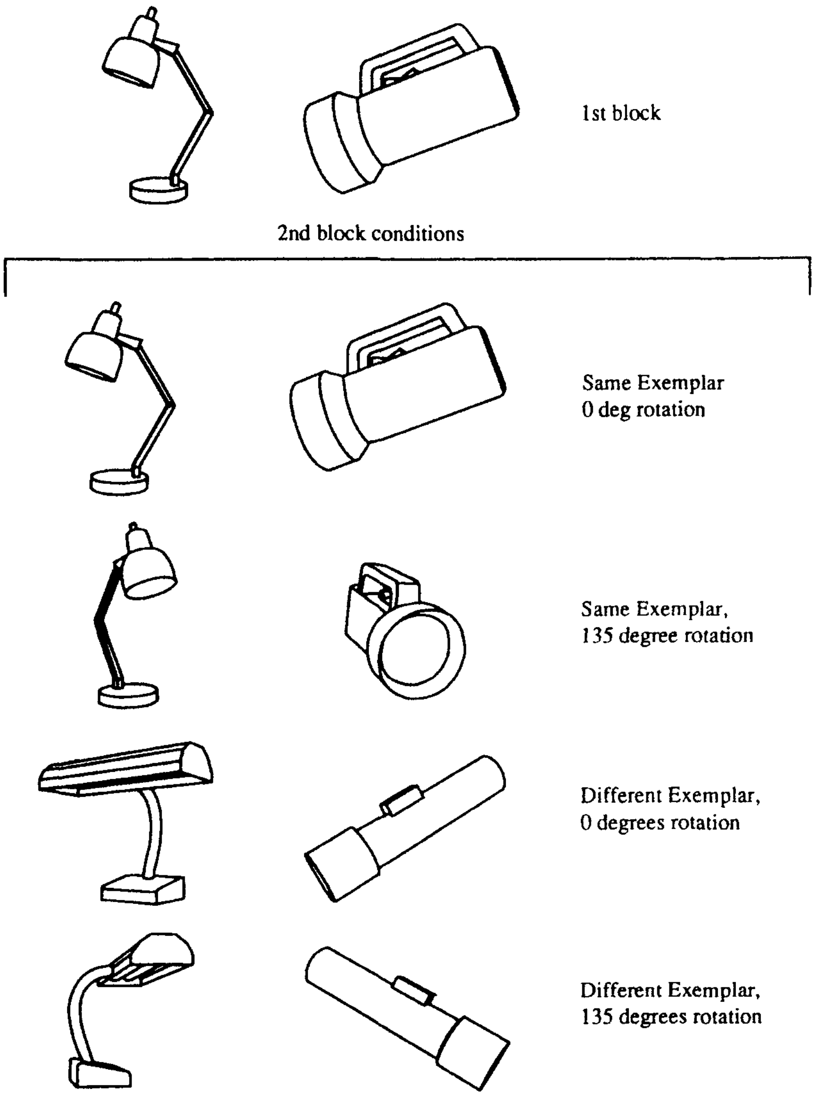

Figure 5. Two sample stimuli illustrating the same and different exemplar conditions, and the 0° and 135° rotation angles of Experiment 1 on name priming of familiar objects.

Stimuli were presented on a Mitsubishi 15-in. (38-cm) monitor (1024 × 768 pixels, vertical refresh rate 60 Hz), controlled by a Macintosh II computer.

Procedure. The task was self-paced in that subjects began each trial by pressing a mouse button. A fixation dot was then presented for 500 ms in the center of the screen, followed by a 200-ms presentation of an object (100 ms in the second block). Following presentation of the object, a pattern of randomly strewn lines (a mask) was presented for 500 ms. Subjects were told that their task was to name the object as quickly and accurately as possible and that they should ignore the mask. Naming RTs were recorded with a Grasen-Stadler voice key attached to a National Instruments timing board (LAB NB-MIO-16H), which afforded timing accuracy to the millisecond. The experiment was run using the Picture Perception Lab software package (Kohlmeyer, 1992). Subjects were provided with overall feedback (mean RT and percent correct responses) at the end of the experiment, as well as response time and accuracy feedback at the end of each trial and trial block. New trials were signaled by a screen request to press the mouse button. Subjects were debriefed as to the purpose of the experiment after their participation.

The experimenter recorded naming errors, false starts (stutters), and subject utterances that failed to activate the voice key, by keying the errors into the computer. A response not made within 3 s was classified as an error. (These were very rare, averaging less than one instance per subject.) Subjects were given 12 practice trials before the experiment began. The primed block followed the priming block immediately in the procedure, with the primed presentation of an object following the priming presentation by approximately 5–7 min and on average 24 trials.

Design. For each object, one of the extreme views (0° or 135°) was arbitrarily designated as Pose A and the other as Pose C with the intermediate view as Pose B. Each subject was shown one exemplar of each object (12 objects in Pose A, 12 objects in Pose C) in a priming block and then was shown either the same exemplar in Pose A, B, or C, or the other exemplar of the object in Pose A, B, or C in a second primed block. This design resulted in three conditions: pose (view A or view C), rotation (orientation change) between first and second block (0°, 67.5°, and 135°), and exemplar (same or different). (Poses A and C for the two exemplars of each entry-level pair were matched as closely as possible, as shown in Figure 5. If the left three-quarter view was Pose A for one object, its different-shaped, same-name exemplar would have as its Pose A an orientation that was as close as possible to a left three-quarter view.) The pose variable was a "dummy" variable included to balance the extreme views in case they differed in canonicality. This did not prove to be the case. The effect of pose (A vs. C) on performance was negligible. Means RTs and error rates for Pose A were 782 ms and 9% errors; for Pose C, they were 801 ms and 12% errors, $t(47) < 1$ for RTs, $t(47) = 1.46, p > .05$, for errors. The results are presented collapsed over the pose variable.

The design was balanced such that each subject saw two objects in each of the 12 cells produced by the combinations of Pose × Rotation × Exemplar conditions. The two exemplars for each object served as prime and target an equal number of times over all views. The design was balanced for order (forward and reverse) such that all the images' mean serial positions (12.5 within each block) were equivalent across conditions and subjects. Thus, 24 pairs of subjects saw different objects in each of the six conditions. Each subject pair saw exactly the same objects in the same conditions but in the opposite presentation orders.

Results

Figure 6 shows the RTs and error rates for Experiment 1. There was sizable priming in that mean correct RTs and errors decreased by 127 ms and 4.8%, respectively, between Blocks 1 and 2, $t(47) = 13.77, p < .0001$, for RTs and $t(47) = 4.24, p < .0001$, for errors. On the second block, RTs for the same exemplar stimuli exhibited a sizable advantage over the different exemplars, by 38 ms, indicating that a portion of the priming was visual, $F(1, 23) = 39.57, p < .0001$. Rotation between prime and target images produced only a small effect on RTs that fell short of significance, $F(2, 46) = 2.74, p = .075$. The Exemplar × Rotation interaction was not significant, $F(2, 46) < 1$. None of the main effects of orientation change, exemplar, or their interaction was close to significant in the error data.

These results indicate that visual priming of naming latency for familiar objects was relatively insensitive to changes in depth orientation occurring between priming and primed images. All views showed an advantage over the different exemplars with the same amount of rotational change, demonstrating that visual (rather than simply object concept or general practice) priming was occurring. Insofar as the advantage of same over different exemplar provides a measure of the experiment's power, the lack of any significant difference between views could not be attributed to the insensitivity of the design.

However, the lack of an effect of rotation could have been the result of a floor effect, as suggested by theories that assume that viewpoint invariance with common objects derives from their familiarity at different views. Because of the familiarity with the object classes (not images), the subjects somehow might have been responding near the naming latency floor over all orientations. That there was sufficient floor to reveal an advantage of same over different exemplars would suggest that this possibility was somewhat implausible. Nonetheless, to test this possibility, each of the two exemplars of each object was classified as hard or easy, according to their mean second-block, same-exemplar RTs. A floor effect would be revealed as an interaction between difficulty and rotation angle, with the hard objects exhibiting an increasing effect of rotation. An ad hoc analysis of variance (ANOVA) was run with object, difficulty, and rotation as the factors on the resulting 498 (out of a possible 576) valid second-block, same-exemplar observations, the remainder being error trials and invalid trials (because of inadvertent tripping of the voice key, etc.). This analysis had spuriously high power because the observations were treated as independent though individual subjects contributed as many as 12 of the 144 Object × Level × View conditions.

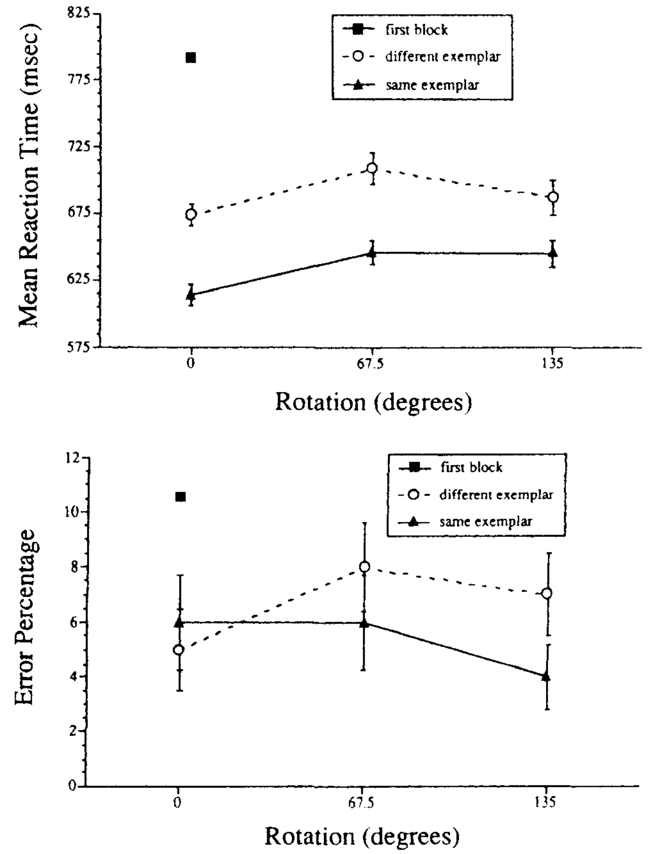

Figure 6. Mean correct reaction times (RTs; *top*) and error rates (*bottom*) on the second block of trials as a function of orientation change (rotation) and exemplar in Experiment 1 on name priming of familiar objects. (Mean RT and error rate on the first block are also shown. The slope of the same-exemplar RT function was 4,355°/s. Because this was a within-subject design, error bars show the standard error of distribution of individual subjects' difference scores, computed by subtracting each subject's mean score on the second block from that subject's score for a particular condition and thus do not include between-subjects variability.)

Figure 7 shows the data for the Difficulty × View interaction. The main effect of difficulty was highly significant,

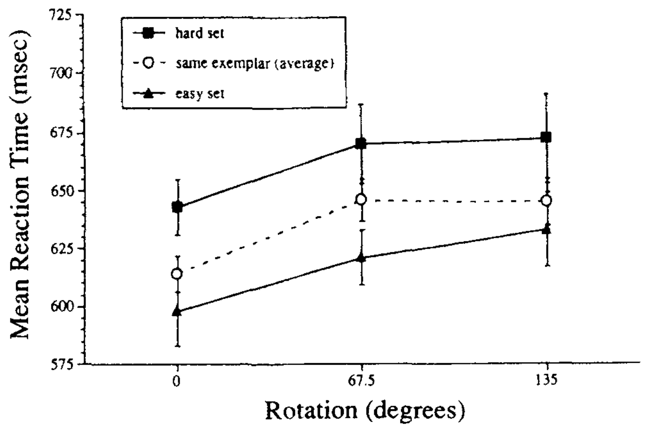

Figure 7. Mean correct naming reaction times for the hard and easy members of each exemplar pair, as determined by a post hoc ranking by mean reaction time for each pair of same-name exemplars, collapsed across object and plotted as a function of orientation change [rotation] from Block 1 in Experiment 1. (Because this was a within-subject design, error bars show the standard error of distribution of individual subjects' difference scores, computed by subtracting each subject's mean score on the second block from that subject's score for a particular condition, and thus do not include between-subjects variability.)

$F(1, 354) = 14.42, p < .0003$, an effect expected from the a posteriori classification of each exemplar as easy or difficult according to its reaction time. The effect of rotation would not have been significant with an appropriate test; however, it managed to reach significance, $F(2, 354) = 3.60$, $p < .03$, because of the spuriously high power. A significant Level × Rotation Angle interaction would be evidence of a floor effect. The interaction, however, like all the others in this analysis, was not significant, $F(2, 354) < 1$, indicating that the hard and easy groups, while differing in absolute level, varied in qualitatively similar ways across views and objects. Thus, there was no evidence that a floor effect was present in the data that could have masked an effect of rotation angle in the original analysis.

At best, there was an extremely small effect of changes in orientation for the same-exemplar condition over the 135° range: The rotation speed based on the best linear fit was 4,355°/s, dramatically faster than that of Tarr's (1989) experiment, where the speed was only 167°/s, after considerable practice. It is difficult to know whether this tiny effect of orientation in the present experiment represented the operation of an orientation-sensitive mechanism or whether it was a consequence of our inability to completely satisfy Condition 3 so the same structural description would characterize all three views of a given object. Even with the negligible effect, the ease of recognizing these particular images at an arbitrary orientation stands in marked contrast to the great difficulty required for the recognition of stimuli, such as those of Edelman and Bülthoff (1992), Rock and DiVita (1987), or Tarr (1989), that failed to satisfy the conditions for invariance. The argument that performance benefited from familiarity can be rejected both on theoretical grounds—none of the pictures were seen by the subjects prior to the experiment—as well as empirical grounds, in that there was no evidence for a floor effect.

Experiment 2: Priming Familiar Objects With and Without Part Changes

Experiment 2 was designed to provide a more direct replication of Bartram's (1974) experiment and to allow a determination of whether rotations that produced changes in the parts present in the image (because of occlusion and accretion of the parts) could have accounted for the effect of orientation changes documented in his experiment. This was done by not adjusting orientations to reduce part changes as was done in Experiment 1. Assuming that these stimuli meet Conditions 1 (readily identifiable parts) and 2 (distinctive GSDs), this experiment provides an indirect test of Condition 3 (identical GSDs).

Method

Subjects. Forty-eight native-English-speaking subjects with normal or corrected-to-normal vision achieved a preset cutoff criterion of correctly naming at least 90% of the stimuli. One subject was excluded by this criterion. Subjects participated for payment or credit in Introductory Psychology at the University of Minnesota.

Stimuli and procedure. The objects were the same as those in Experiment 1, with the exception that the initial poses differed for many of the objects and the amount of orientation change between views was halved. Thus, Poses B and C differed by 33.25° and 67.5°, respectively, from Pose A. Poses A and C were not necessarily the same as the 0° and 67.5° views used in Experiment 1. (As in Experiment 1, there was no reliable effect of prime—A vs. C—on first block RTs, or errors—View A = 706 ms and 3%, View C = 721 ms and 3% $t(47) = 1.23, p > .01$, for RTs and $t(47) < 1$ for errors—so each subject's data were collapsed across the priming view for all subsequent analyses.) No attempt was made to maximize the overlap in the parts among the three views for a given object, as was done in Experiment 1. The procedure and design were the same as those in Experiment 1.

Results

The RTs and error rates for Experiment 2 are shown in Figure 8. There was considerable priming in that second block RTs were lower than those in Block 1 by 75 ms, $t(47) = 10.29, p < .0001$, but the errors rates were lower by only 1%, $t(47) = 1.30$, *ns*.

Mean correct RTS and error rates to the same exemplars on the second block were lower by 24 ms, $F(1, 23) = 8.96$, $p < .01$, indicating, as in Experiment 1, that a portion of the priming was visual. There was only a 0.6% advantage in error rates for same over different exemplars, $F(1, 23) < 1$. Although there was an apparent slight increase in RTs with increasing disparity between priming and primed orientations, the Exemplar × Rotation interaction was not significant, $F(2, 46) = 1.41, p > .20$. The only test with the errors that was close to significance was the Exemplar × Rotation interaction, $F(2, 46) < 2.13, p > .13$.

Though the Exemplar × Rotation interaction fell short of significance, there was, nonetheless, a sufficient increase in the same exemplar condition over rotation that at a rotation of 67.5°, performance was close to that of the different exemplar condition and would not have differed statistically from it. To evaluate whether changes in the parts could be responsible for this effect, three judges each familiar with geon theory, rated the pairs of Poses A–B, A–C, and B–C for each of the 48 objects. The raters were asked to judge whether or not the two had equal numbers and types of parts clearly visible. Pairs given two "no" answers were removed from the data set for purposes of calculating the results for an adjusted same-exemplar condition at each orientation. The number of objects removed was: 0 at View A, 9 at View B, and 22 at View C. Figure 9 gives an example of a pair in which the C pose was removed.

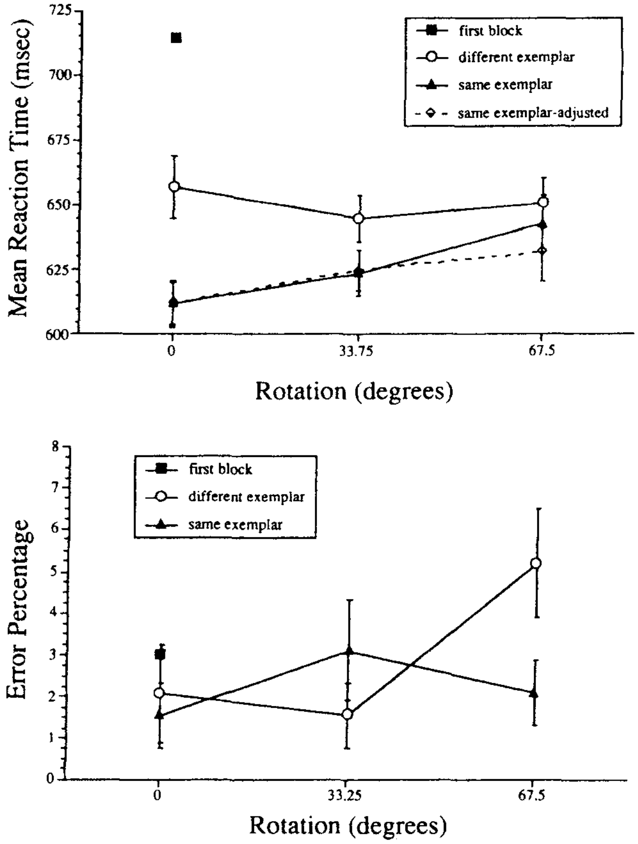

Figure 8. Mean correct reaction time (RTs; *top*) and error rates (*bottom*) on the second block of trials as a function of orientation change (rotation) and exemplar in Experiment 2 on name priming of familiar objects. (Mean RT and error rate on the first block are also shown. The slope for the RTs from the same-exemplar condition for all the objects presented in the experiment [solid triangular points labeled "same exemplar"] was 2,177°/s. The slope for the RTs for only those objects determined by a post hoc examination to contain the same parts across the change in orientation [open diamond points labeled "same exemplar-adjusted"] was 3,376°/s. Because this was a within-subject design, error bars show the standard error of distribution of individual subjects' difference scores, computed by subtracting each subject's mean score on the second block from that subject's score for a particular condition, and thus do not include between-subject variability.)

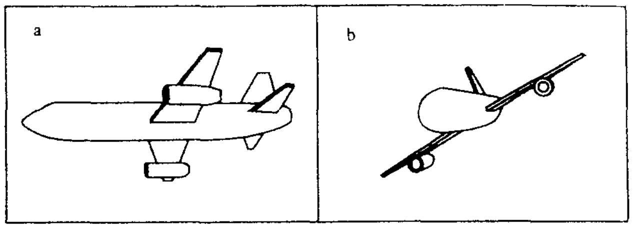

Figure 9. An example of an image pair that was omitted from the "same-exemplar-adjusted" function in Figure 8. (The number of parts differs in the two views of the airplane [9 on the left and 6 on the right]. The volumetric shape that can be inferred from the contour of many of the parts also differs. For example, the engine on the left wing in the left panel is identifiable as a cylinder but appears as an annulus in the right panel.)

Removing these objects produced the line labeled same exemplar–adjusted in Figure 8. It is clear that this function reveals a smaller effect of rotation than that for the original data. The slope of this function was 2,177°/s prior to removing those objects with parts that changed from priming to primed view and 3,376°/s after those objects were removed.

It is thus possible that part changes (i.e., failure to meet Condition 3) were responsible for the disadvantage of Bartram's (1974) different view condition, relative to his identical condition. A similar failure to meet Condition 3 likely accounts for the effects of orientation differences in priming in a recent report by Srinivas (1993). Selection of different orientations of complex objects without regard to how they might differ in the parts (or any other information) that are present in each pose will very likely produce a cost of orientation disparity (from studied to tested views) as parts are occluded or revealed. Srinivas's failure to specify what did or did not change from one view to another thus renders her results of limited relevance to current accounts of shape representation.

Experiment 3: Recognition of Nonsense Objects With and Without Part Changes

Experiments 1 and 2 provided evidence that human object recognition is largely invariant to large changes in depth orientation when the stimuli and changes satisfy the conditions for invariance. Those experiments were performed with images that depicted instances of familiar classes. Experiment 3 was designed to determine whether viewpoint invariance could be immediately obtained with unfamiliar objects as long as they satisfied the conditions of invariance.

The stimulus set (shown in Figure 10) was composed of objects that could readily be described in terms of a distinctive GSD, thus meeting Conditions 1 and 2 for invariance.

A sequential, same-different matching task was used in which a depicted object could undergo two types of rotations in depth between its first and second presentations on a same trial. One rotation condition (no parts change condition) met Condition 3 in that the same GSD would be activated by both

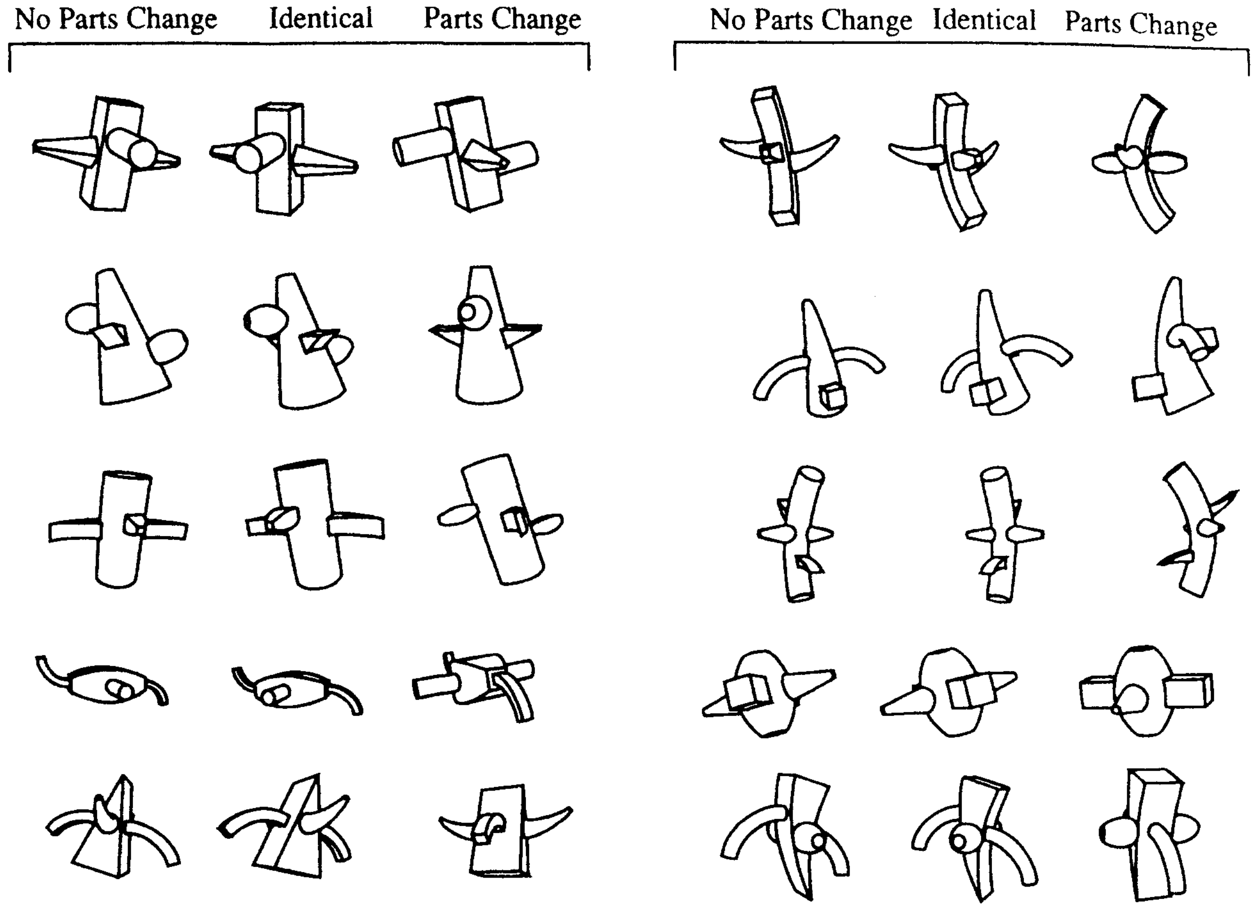

Figure 10. The 10 unfamiliar objects used in Experiment 3. (The no-parts-change and parts-change views are rotations of 45° in depth in different directions from the zero view. Note that no object contains a part unique to that object, and relations between object parts are the same for all objects.)

images. In the other rotation condition (parts-change condition), the rotation largely occluded one small part (of a five-part object) and revealed another (that had been largely occluded in the first view) so as to entail a GSD differing in one geon from its initial pose so as to not satisfy Condition 3. The smaller geons at opposite sides of the large geon were identical so the geon that was revealed in the parts-change condition was predictable (and often could be discerned when it was partially occluded). The objects were comprised of a sample of the same set of geons, and all had a large central volume with smaller volumes spaced around it at approximately 90° intervals. No single geon or relation, therefore, was sufficient to distinguish among the objects: A particular geon in a particular relation had to be specified for an object to be uniquely identified. For example, the largest (central) geon was sufficient to distinguish among the stimuli. However, both size (larger than) and type of geon had to be specified.

Method

Subjects. Sixteen subjects with normal or corrected-to-normal vision achieved a preset cutoff criterion of 80% overall accuracy in the experiment. No subjects were excluded by this criterion. Subjects participated for payment or credit in Introductory Psychology at the University of Southern California.

Stimuli. Ten nonsense objects were drawn in the same manner as the familiar objects used in Experiments 1 and 2. Each object contained 5 of a set of 10 parts, each describable as a simple three-dimensional volume corresponding to a geon class as defined by geon theory. (The 10 parts were straight and curved major-axis versions of a brick, a cylinder, a truncated cone, and a wedge, and a straight cross section and a curved cross-section version of a lemon, as shown in Figure 10.) Each geon was the central part of one object, and pairs of two of the remaining nine parts were attached as peripheral parts to the front, back, left, and right of the central part such that the two members of a pair were opposite each other. Three views of each object were chosen such that (a) two were 45° rotations in opposite directions in depth of a central pose and (b) the central pose and one of the other poses contained exactly the same parts (although not from the same orientation and not the same local features) to define a no-parts-change condition, while the other pose revealed a part that was largely occluded in the other two views but occluded a part that was visible in the other poses, to define a parts-change condition.

Procedure. The experiment was run as a sequential matching task with a relatively brief interstimulus level (ISI). Subjects were told that when the two images were of the same object, the second image would sometimes be from a different orientation in depth than the first. Prior to the experimental trials, they were shown the set of stimuli printed on a sheet of paper and made aware of the structure of the stimulus set, namely, that rotation of an object might result in a change in one of the small parts. They were instructed to ignore such changes, and were given 12 practice trials (with a set of objects not used in the experiment) in which changes of both types present in the experimental images occurred.

As in the prior experiments, subjects began each trial by pressing a mouse button. A fixation dot was then presented for 500 ms in the center of the screen, followed in turn by a 200-ms presentation of an object and a 750-ms presentation of a mask.[10] A second object image was then presented for 100 ms, followed by a second mask for 500 ms. Subjects were instructed to ignore the intervening mask, and when the second image appeared, to press a microswitch key as quickly as possible if the second image was of the same object as the first. They were told that if the two images were of different objects, they were to do nothing (a go-no-go task). (With very easy perceptual tasks, much of the variability in RTs is a consequence of response selection. The go-no-go task was chosen to reduce this variability in Experiments 3, 4, and 5.) Overall feedback (mean RT and percent correct responses) was provided after 60 trials and at the end of the experiment, as was response time and accuracy feedback at the end of each trial, followed by a prompt for the next trial.

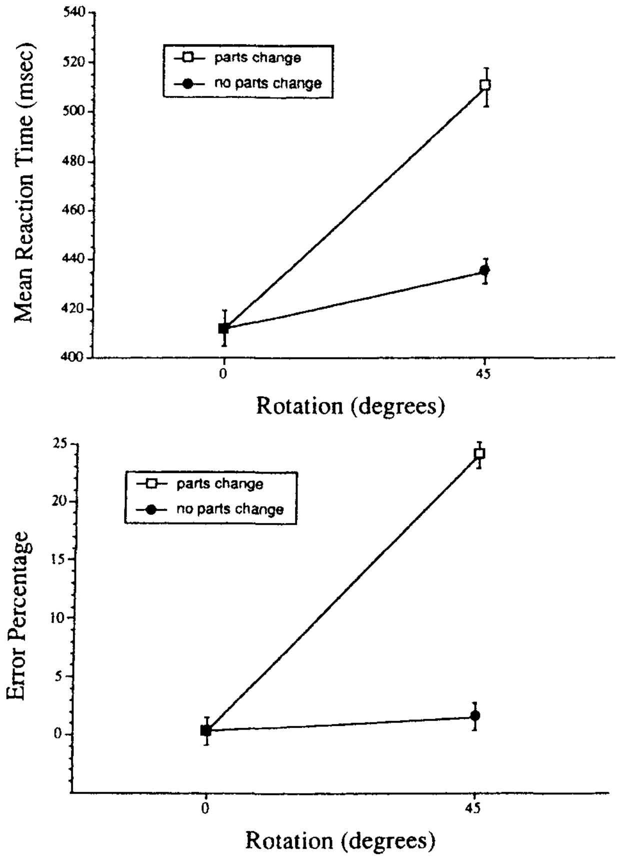

Figure 11. Mean correct reaction times (*top*) and error rates (*bottom*) for same–different judgments of unfamiliar objects in Experiment 3 as a function of the degree of angular change between the first and second exposures on a trial and indication of whether that rotation produced a part change. (The speed for the no-parts-change reaction time function was 1,875°/s. For the parts-change function, the speed was 459°/s. Because this was a within-subject design, error bars show the standard error of distribution of individual subjects' difference scores, computed by subtracting each subject's mean score from that subject's score for a particular condition, and thus do not include between-subjects variability.)

Design. Each subject performed 120 trials. The first image shown in each trial was one of the three views of an object. The second exposure was always the B pose of 1 of the 10 objects (the same object on positive trials). Thus, three orientation conditions were defined: an identical condition (Pose B followed by Pose B), a no-parts-change condition (Pose A followed by Pose B), and a parts-change condition (Pose C followed by Pose B). Each subject saw each object once in each of the three conditions in one positive and one negative trial in each of the two blocks of the experiment, resulting in 12 combinations of conditions: 3 (part-change condition: identical vs. parts change vs. no parts change) × 2 (block) × 2 (response), with each object appearing once in each of the 12 cells. (As a go-no-go task was used, the response variable could only be analyzed for the error rates.) Order of trials within blocks was randomized separately for each block. The design was balanced for order such that all objects' mean serial presentation position was equivalent within and across blocks and across subjects. Four subjects were run in each order.

Results

Individual observations were averaged across objects for each orientation condition, resulting in a total of six data points per subject. Mean correct RTs and error rates on positive trials are shown in Figure 11. There was virtually no effect of rotation unless the rotation produced a parts change. RTs and errors were analyzed with a repeated-measures ANOVA on block and rotation condition (identical, no parts change, or parts change) as fixed variables. (The error ANOVA also included answer as a variable.) For the RTs, the effect of rotation condition, $F(2, 30) = 39.30, p < .001$, was highly significant, with almost all of the effect attributable to the increased RTs in the part-change condition. Bonferroni t tests (which adjust alpha levels when multiple t tests are performed) indicated that the difference between the parts-change and identical conditions was significant, $t(45) = 3.96, p < .005$, as was the difference between the parts-change and no-parts-change conditions, $t(45) = 2.99, p < .05$, but was not significant between the no-parts-change and identical conditions, $t(45) = <1.00$. Neither the effect of

[10] The 750-ms ISI was selected on the basis of an experiment by Ellis and Allport (1986). These investigators showed that at brief (100-ms) ISIs, same–different RTS were affected by physical differences in size, position, or orientation between the images. At 750 ms, the matching was invariant. Presumably, at 100 ms subjects were able to match the second stimulus against iconic activity from the first image. By 750 ms, this activity was no longer available and the subjects matched against a longer lasting memory representation.

block, $F(1, 15) = 2.86$, nor the Block × Parts Change interaction, $F(2, 30) < 1$, was significant.

As with the RTs, the effect of rotation on error rates for same trials was highly significant, $F(2, 30) = 18.09$, $p < .001$, with all the effect attributable to the increased error rates in the parts-change condition as indicated by the Bonferroni tests. The difference between the identical and parts-change conditions was significant, $t(90) = 8.02$, $p < .001$, as was the difference between the no-parts-change and parts-change conditions $t(90) = 7.59$, $p < .001$, but the difference between the no-change and identical conditions was not significant, $t(90) < 1.00$, *ns*. There was a lower miss rate (38.8%) than false alarm rate (14.2%), leading to a significant effect of answer, $F(1, 15) = 6.50$, $p < .001$. On the different trials, the false alarm rate (25.0%) for the identical condition was much higher than either of the other two conditions (both were 8.8%), leading to a significant Rotation × Answer interaction, $F(2, 30) = 56.49$, $p < .001$.

The rotation speed between the zero condition and the no-parts-change condition was 1,875°/s, and the speed between the zero condition and the parts change condition was 459°/s. Note that the speed of the depth rotation without a parts-change found in this study is more than an order of magnitude faster than the initial speed reported by Tarr (1989; 167°/s), and almost four times as fast as the speed reported by Edelman and Bülthoff (1992; 514°/s) after large amounts of practice in a comparable task. Recall that the slope from Tarr's (1989) investigation improved to about the same level as that of Edelman and Bülthoff after more than 100 trials with each image, a rate substantially slower than the initial no-parts-change slope in this experiment.

Why was the speed of the no-parts-change condition, as fast as it was, still slower than the speeds in Experiments 1 and 2? Greater similarity among the objects used in this experiment may have magnified the effects of rotation in depth. To obtain the specific stimulus orientation differences required to test the hypothesis in this experiment, it was necessary to create a set of highly regular objects. Each of the 10 objects had exactly the same number of parts, in the identical relations, and (with one exception) the same orientation of the central part and the same orientation of all the peripheral parts. Thus, while the members of this stimulus set could be distinguished by the geons, all the other attributes were virtually identical across these objects. The objects would thus have had small Hamming distances. In contrast, a relatively arbitrary assemblage of objects depicting different entry-level classes, as was used in Experiments 1 and 2, will have relatively large Hamming distances in that they will differ in all possible attributes and thus allow more rapid differentiation at arbitrary orientations.

It is of interest to note that a change in a small part led to reliable increases in RTs and error rates even though the small parts could have been ignored in performing the task. That the subjects did not (or could not) simply use the largest central geon to perform this task argues against the invariance being a consequence of a simple feature search. This result is considered in the General Discussion section.

Experiment 4: Matching of Single Volumes

Experiment 4 tested whether depth rotation effects would be present in the matching of single volumes that differed from the distractors in terms of geon classes.

A remarkable phenomenon about mental rotation studies is that a single exposure of a test stimulus leads to long-lasting orientation effects such that recognition is fastest at the original orientation and progressively slower at greater angular disparities (e.g., Shepard & Cooper, 1983). If this phenomenon is present when distinguishing among geons (rather than determining a stimulus' left–right orientation, as in the mental rotation studies) then a cast of rotation should be evidenced.

Method

Subjects. Twenty native-English-speaking persons (12 female, 8 male; 18–26 years of age) from the University of Southern California participated for payment or course credit in Introductory Psychology. All subjects had normal to corrected-to-normal vision.

Stimuli. Ten volumes (brick, curved brick, truncated cone, curved truncated cone, cylinder, curved cylinder, wedge, curved wedge, curved cross-section lemon, straight cross-section lemon) were constructed in a three-dimensional drawing package (Swivel Professional, Paracomp, San Francisco). Three poses of each object, differing by 45° and corresponding to three views about the vertical axis in a right-handed world coordinate system, were exported from this package and redrawn in Adobe Illustrator (Adobe Systems, Mountain View, California) to produce line drawings of each of the views. Figure 12 shows the three poses of each volume used in Experiment 4. Poses were chosen so as to minimize accidental views,[11] but this attempt was not completely successful because some of the poses presented little or no three-dimensional vertex information (i.e., forks, Ys, or tangent Ys).

The different volumes shared many individual lines and vertices, such as a curved line or a tangent-Y vertex, or individual volume descriptors, such as a curved axis, which precluded their use as simple features for performing the task.

The pose of each volume that presented the largest span (in any single two-dimensional direction) was scaled to fit a circle the diameter of which subtended a visual angle of 4.5°. The other poses of each object were scaled by the same amount as the largest view so as to preserve their size relative to that view. Stimuli were presented on a 16-in. (41-cm) Sony Trinitron monitor attached to an E-Machines video board (Model TX, 1,024 × 768 pixels, vertical refresh rate 70 Hz) controlled by a Macintosh IIfx computer.

Procedure. The procedure was the same as that used in Experiment 1 with the following exception. A target-learning trial occurred first in each block. On this trial, the subject saw a volume for 20 s. Subjects were instructed to look at the target and remember it; they were told that they would have to identify the target in a set of test trials following the learning trial. After the learning trial, 18 test trials were presented. In each test trial, a fixation dot was presented for 500 ms in the center of the screen, followed by a 150-ms presentation of a volume. Following the presentation of the volume, a mask of random lines was presented for 500 ms. Subjects pressed a microswitch key if the volume that appeared was the target. They

[11] In the present case, accidental views are those from which the identity of the geon, its aspect ratio, or both could not be readily determined.

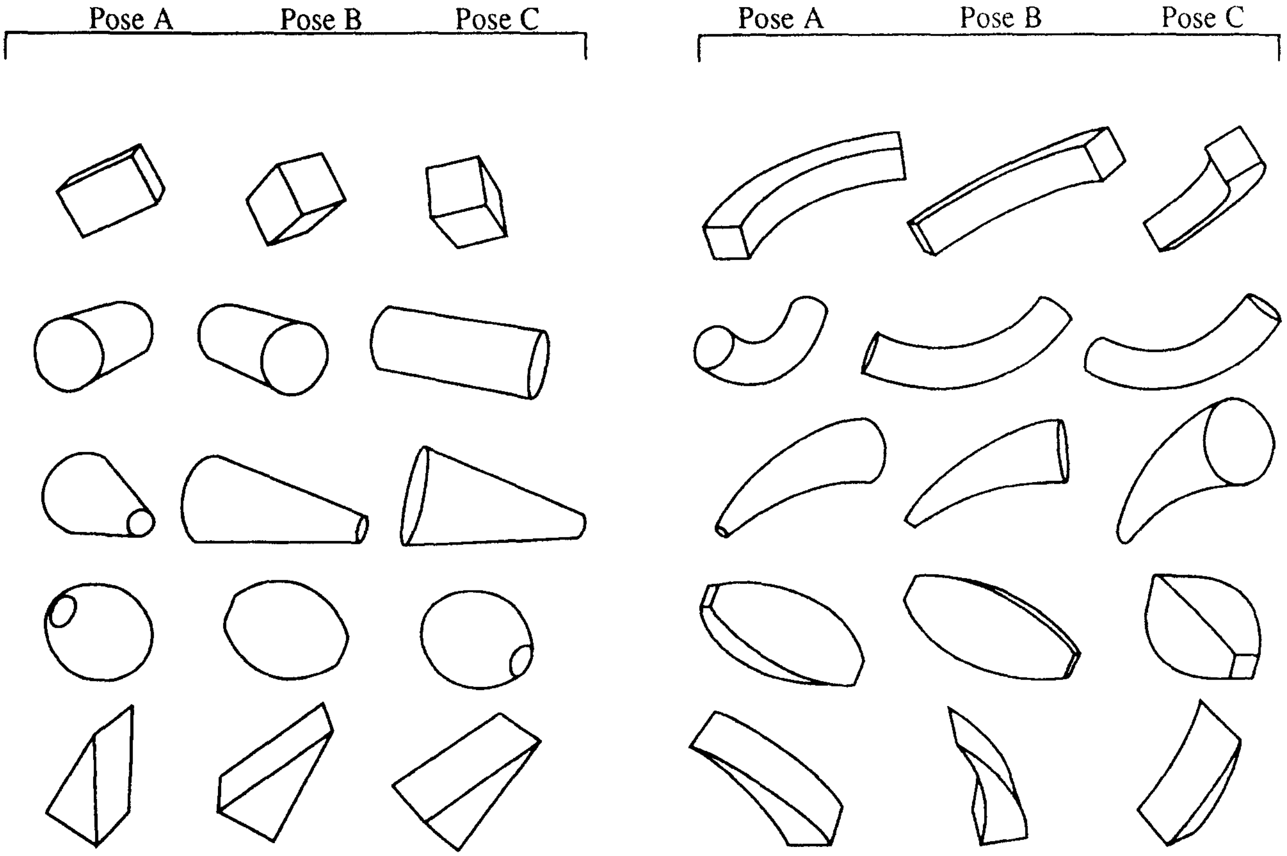

Figure 12. The 10 volumes (geons) of Experiment 4, in each of the presented orientations. The orientations differ by 45° in depth. (Note that some views are nearly accidental in that they contain almost no three-dimensional vertex information [forks, arrows]. The middle view of the curved cross-section lemon is an example.)

were to do nothing if the volume was not the target. Subjects were told that the target volume would appear in both the learned orientation and in new orientations, but that they were to ignore the orientation changes as best as they could. After the 18 test trials, the subject was presented with a new target. This pattern continued until the subject was tested on all of the 10 volumes used in the experiment.

If the subject failed to respond within 1.5 s, the trial was counted as an error. Subjects were given a practice cycle of 1 learning and 12 test trials—with nonsense stimuli not present in the experimental trials—before the experiment began.

Design. For each volume, one of the extreme views (0° or 90°) was arbitrarily designated as Pose A and the other as Pose C, with the intermediate view designated Pose B. Each subject was shown one pose of one volume in a learning trial. Each of the three poses (A, B, and C) appeared three times in the test trials, and every other volume in the set appeared once in the set of 18 test trials (as a distracter). This conformed to a 2 × 3 design: pose (A vs. C) and orientation change (0°, 45°, and 90° from the learned pose). Pose was included in the ANOVA to balance the extreme views in case they differed in canonicality. (As in the previous experiments, this did not prove to be the case; the effect of pose on naming latency was negligible—Pose A, 292 ms and 5% errors, Pose C, 303 ms and 4% errors, $t(19) = 1.32, p > .05$, for RTs; $t(19) = 1.59, p > .05$, for same errors—so the results are presented collapsed over this variable.)

The design was balanced for order (forward and reverse) such that the mean serial positions of all of the volumes (9.5 within each test block) were equivalent across conditions and subjects. Ten pairs of subjects learned the same target orientations, each in a different order, determined by a Latin square. Each subject pair saw exactly the same objects in the same conditions, but in opposite presentation order.

Results

There was virtually no effect of orientation changes. Figure 13 shows the mean correct RTs and error rates for positive trials for Experiment 4. The mean RTs were 296, 304, and 293 ms for the 0°, 45°, and 90° orientation changes, respectively. A repeated-measures ANOVA on the RTs showed a barely significant effect of rotation (0°, 45°, 90°), $F(2, 38) = 4.20, p < .05$, although this was a nonmonotonic effect caused by slightly higher RTs in the intermediate (45°) rotation condition than in either the 0° or 90° rotation conditions. (RTs to 90° rotations were actually slightly lower than RTs at 0° rotations.)

Mean error rates actually declined with increasing rotations. For the 0°, 45°, and 90° rotations, error rates were

5.0%, 4.3%, and 1.7% for positive trials and 20.3%, 16.3%, and 16.7% on negative trials. However, the effect of rotation on the error rates fell short of significance, $F(2, 38) = 1.78$, *ns*. As in the prior experiment, the false alarm rate was higher than the miss rate (17.8% to 3.7%), leading to a significant effect of answer; $F(1, 19) = 33.48$, $p < .001$. The interaction between rotation and answer was not significant, $F(2, 38) < 1.00$.

Recall that no volume in the set contained a unique vertex, line, or volumetric feature that could serve as a cue for its recognition. Thus, the results of this experiment provide evidence for strong, immediate invariance over changes in the depth orientation in the recognition of single volumes.

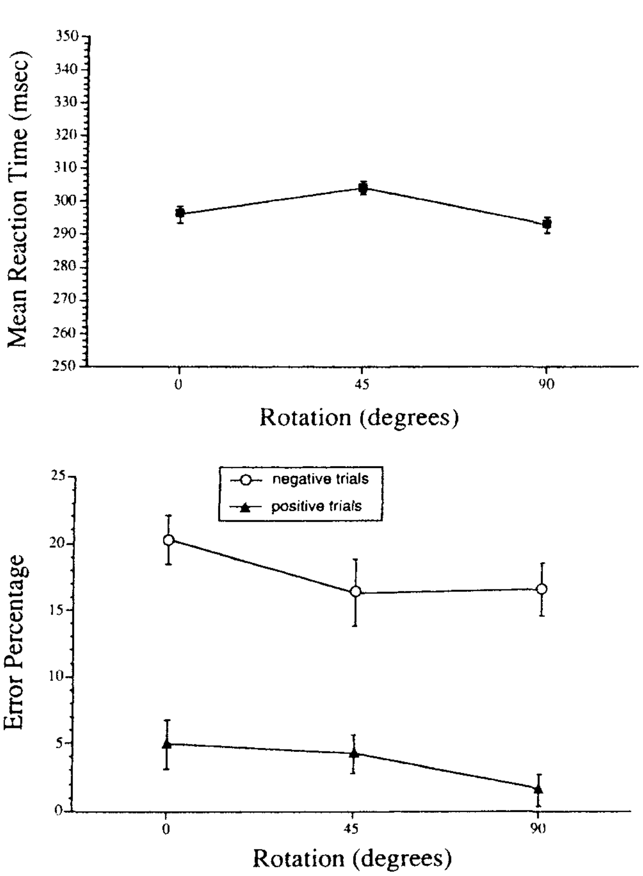

Figure 13. Mean correct reaction times (RTs; *top*) and error rates (*bottom*) for detecting individual geons (Experiment 4) as a function of amount of rotation relative to the learned orientation. (As this was a go–no-go task, only the "yes" responses for the RTs were obtained. For the error rates, negative trial orientations were determined relative to the orientation [A or C; see Results section under Experiment 4 of text] of the studied view of the target volume. The RT slope was slightly negative. Because this was a within-subject design, error bars show the standard error of distribution of individual subjects' difference scores, computed by subtracting each subject's mean score from that subject's score for a particular condition, and thus do not include between-subjects variability.)

Experiment 5: Adding a Distinctive Geon to Viewpoint-Dependent Objects

Experiments 1, 2, and 3 provided evidence that human object recognition is largely invariant to large changes in depth orientation, when the stimuli and changes do not violate the conditions for invariance. These conditions are based on the assumption that geon differences are sufficient for depth invariance. Experiment 4 documented depth invariance for single geons. Experiment 5 was designed to determine whether the addition of a single distinctive geon would be sufficient to confer immediate viewpoint invariance to complex objects that are otherwise highly viewpoint dependent.

As described previously, the subjects in the Edelman et al. (1989) experiment evidenced enormous difficulty in distinguishing among 10 bent "paper clip objects" when these objects were viewed at a new orientation. Each object was composed of five "wires" (or elongated cylinders). The objects differed only in the angle formed at adjacent wires. Our Experiment 5 was a replication of Edelman et al.'s experiment with one small modification: Each of the stimuli had a different geon for its central segment.

Method

Subjects. Twenty members of the University community (12 men and 8 women, ages 20–40 years) with normal or corrected-to-normal vision from the University of Southern California participated for credit or payment.

Stimuli and procedure. Ten objects were constructed in a manner similar to that described by Edelman et al. (1993) with the drawing procedures described for Experiment 4. Five cylinders were connected end-to-end, and the joins between cylinders were perturbed by a random amount in 3D. The set of objects differed from those used by Edelman et al. in that the middle cylinder was replaced by 1 of 10 volumes (brick, curved brick, truncated cone, curved truncated cone, cylinder, curved cylinder, wedge, curved wedge, curved cross-section lemon, straight cross-section lemon) so as to produce 10 objects that differed by one part and by the metric angle of connection between all of the parts, as illustrated in Figure 14. Three views of each object, differing by 30° (±5°) between views were created as described in Experiment 2. These views, which are referred to as Poses A, B, and C, were chosen to match those of Edelman et al. (1989). Because it proved to be impossible to select orientations that did not project accidental views of at least one part for many of the views, the criterion for choosing views was simply to minimize such accidents. The stimuli were scaled to fit a circle spanning 5° of visual angle and were presented with the procedure described for Experiment 3. If the subject failed to respond within 1.5 s, the trial was counted as an error. (Such errors were rare; the mean was 2.5%; the median, 1.5%.) Subjects were given a practice cycle of 1 learning and 12 test trials before the experiment began.

Design. For each object, one of the extreme views (0° or 60°) was arbitrarily designated as A and the other as C, with the intermediate view designated as B. Each subject was shown one exemplar of each object in the learning trial. Each of the three views A, B, and C appeared 3 times in the test trials, and each of the remaining 9 objects in the set appeared once (as a distractor) in the set of 18 test trials. This design resulted in two conditions: pose (A or C), and orientation change (0°, 30°, and 60° from the learned

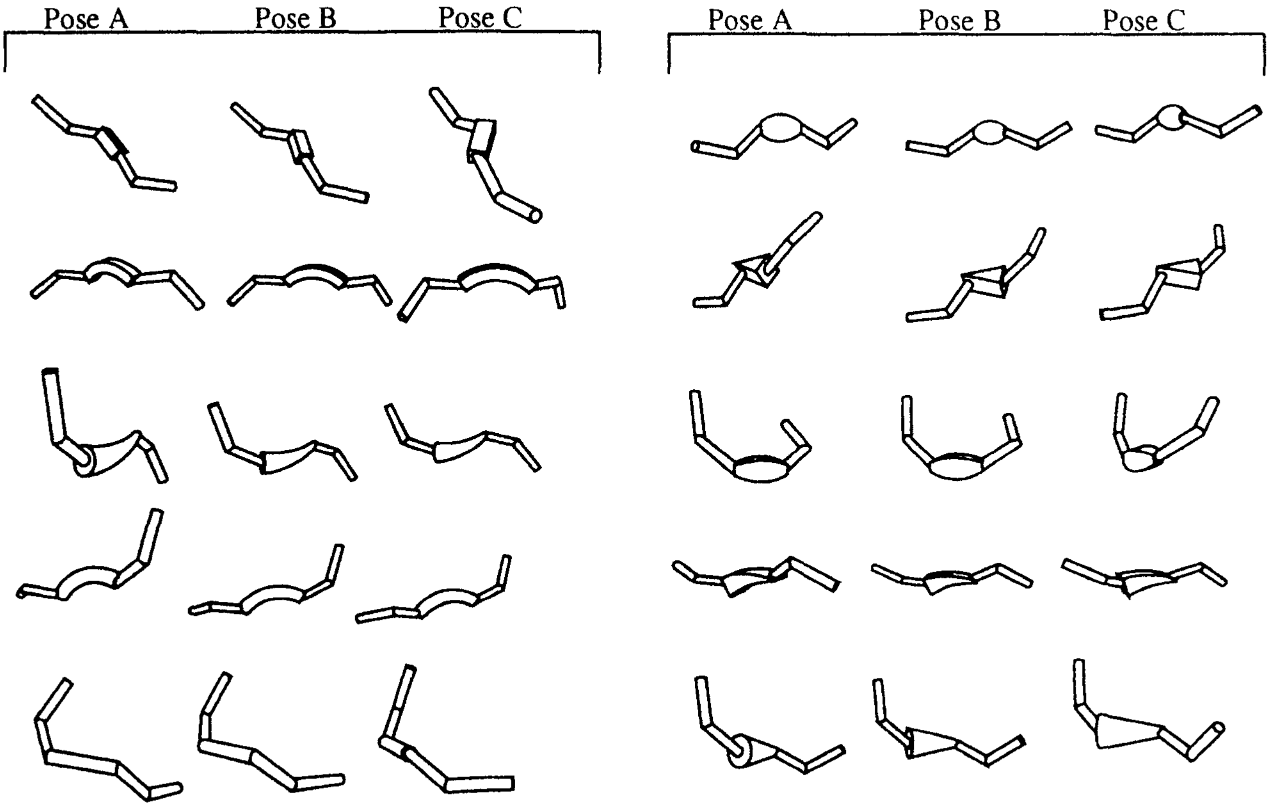

Figure 14. The three poses (each differing by 30° from its nearest neighbor) of the stimuli used in Experiment 5 on rendering viewpoint-dependent objects viewpoint invariant (converting bent paper clips into charm bracelets). (These objects have been redrawn from Figure 2 of Edelman et al. 1989, and have been modified by substituting a distinctive geon for one of the wires in the middle segment for each object. From *Stimulus Familiarity Determines Recognition Strategy for Novel 3D Objects* [Artificial Intelligence Laboratory Tech. Rep. No. 1138, p. 3], by S. Edelman, H. Bülthoff, and D. Weinshall, 1989, Cambridge, MA: MIT Press. Copyright 1989 by S. Edelman. Adapted by permission.)

pose). Learning orientation was included to balance the extreme views in case they differed in canonicity. (This did not prove to be the case; there was only a small effect of priming view—A vs. C—on naming latency: Pose A, 430 ms and 7% errors; Pose C, 397 ms and 4% errors, $t[19] = 1.28$ for RTs, $t[19] = 1.19$ for errors, $p > .10$ for both; thus, the results are presented collapsed over this variable.) The design was counterbalanced for presentation order as in Experiment 3.

Results

Figure 15 shows the mean correct RTs and error rates for positive trials for Experiment 5. The three orientation changes showed mean RTs and error rates of 396 ms, 377 ms, and 408 ms, and 3.1%, 4.6%, and 7.6%, at 0°, 30°, and 60°, respectively, for positive trials. The effect of rotation on RTs was reliable, $F(2, 38) = 17.90, p < .001$, primarily caused by the small *reduction* in RTs at the intermediate orientation. Planned comparisons between orientation changes 0°–30° and 0°–60° indicated that although the 0°–30° change was significant, $F(1, 19) = 17.37, p < .001$, the 0°–60° change was not, $F(1, 19) = 4.30, p > .05$.

The small increase in error rates as a function of rotation was reliable, $F(2, 38) = 6.37, p < .005$, although none of the Bonferroni tests were significant. As in Experiments 3 and 4, the error rate on positive trials was lower than on negative trials, 5.2% versus 10.7%, leading to a significant effect of answer, $F(1, 19) = 10.22, p < .005$. The interaction was not significant.

The slope of the best linear fit to the mean RT at 0°, 30°, and 60°, was 5,000°/s, faster than that of any of the previous experiments with the exception of Experiment 3, which used single volumes as stimuli. Assuming that a rotation function was present in the data, the rate of rotation in this experiment would be an order of magnitude faster than the function reported by Edelman et al. (1993) after extensive practice. The results of the planned comparisons on the RTs, however, argue for an interpretation of complete invariance (for the RTs) with respect to changes in the depth orientation of the target. The only reliable comparison was the 0°–30° contrast, indicating that the intermediate orientation was recognized faster than the studied orientation (396 ms mean RT at 0° orientation vs. 376 ms at 30°). It must be emphasized that the

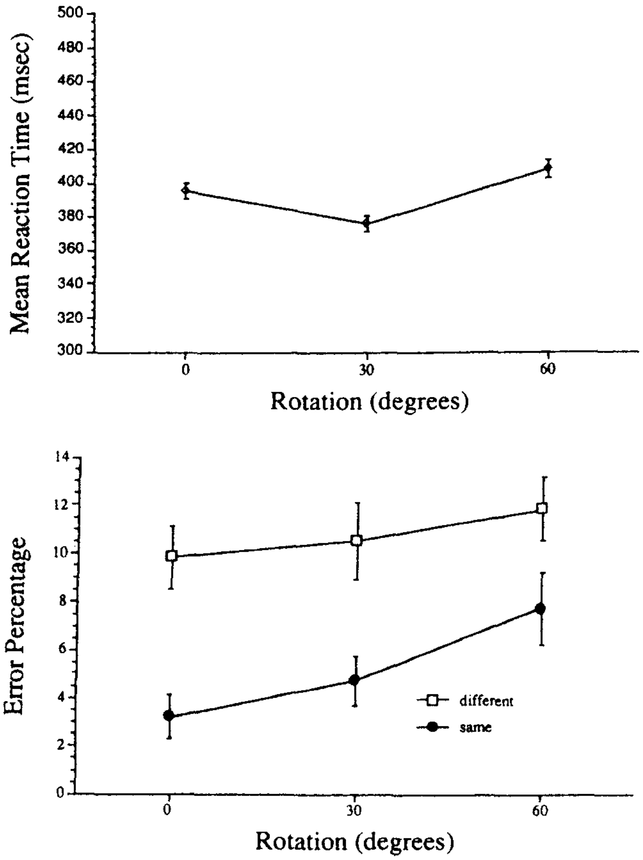

Figure 15. Mean correct reaction times (RTs; *top*) and error rates (*bottom*) in Experiment 5 for same–different judgments of two nonsense objects (the charm bracelets) as a function of rotation angle relative to the studied target orientation. (The RT rotation rate was 5,000°/s. As this was a go–no-go task, only "yes" RTs were obtained. Because this was a within-subject design, error bars show the standard error of distribution of individual subjects' difference scores, computed by subtracting each subject's mean score from that subject's score for a particular condition, and thus do not include between-subjects variability.)

subjects in this experiment achieved near viewpoint invariance in their RTs after viewing the target for only 20 s at a single orientation. In fact, probably a much briefer familiarization exposure would have been sufficient in that the subjects would indicate that they were ready to start the block long before the 20-s familiarization interval was completed. Adding a single distinguishing geon dramatically reduced the extraordinary difficulty present in the Edelman et al. (1993) experiments in recognizing these objects at new orientations. Given the power of the experiment, the modest increase in error rates was reliable. Possible reasons for this are considered in the General Discussion section.

Poggio and Vetter (1992) recently showed that if a three-dimensional object is bilaterally symmetrical (as were most of the objects in Experiments 1–4 in the present investigation) one nonaccidental two-dimensional view is sufficient to achieve recognition of that object from all viewpoints. In the present experiment (Experiment 5), the objects were not symmetrical under any axis, yet viewpoint invariance was achieved from a single view. This result suggests that a distinctive GSD is sufficient for viewpoint invariance for nonsymmetrical objects. Thus, viewpoint invariance can be achieved in the absence of symmetry. Nonetheless, there may be a contribution of symmetry in that Vetter, Poggio, and Bülthoff (1993) reported an advantage for the recognition of symmetrical wire-frame objects over nonsymmetrical versions.

General Discussion

Experiment 1 showed that large rotations in depth had virtually no effect on naming latencies of familiar objects from different entry-level categories, presumably because distinctive geon structural descriptions could be activated from each image. Experiment 2 replicated this result but showed that the invariance was true only when the orientation change was not accompanied by a change in visible parts. Experiment 3 established that orientation invariance could be achieved immediately with the matching of unfamiliar objects that met Conditions 1 and 2 (distinctive GSDs), as long as the rotation was not accompanied by a parts change that would have changed the GSD and thus violated condition 3 (identical GSDs over different viewpoints).

The conditions for invariance were based on the assumption that the detection of individual geons would be viewpoint invariant. This assumption was supported by the results of Experiment 4. The results of several experiments indicate that a difference in a single geon in a complex object was sufficient (a) to confer distinctiveness (near viewpoint invariance) to unfamiliar, originally viewpoint-dependent objects (the charm bracelets of Experiment 5) so as to satisfy Condition 2; or (b) lose invariance when rotation in depth changed a single geon in Experiment 3 (nonsense object matching). Curiously, in Experiment 3, the subjects could have ignored the smaller geons that were changing in favor of the largest geon, which never changed. That they could not do this suggests that the registration of geons is obligatory and a change in a geon produces a change in the structural description of the representation of the object. In contrast, it may be that variation in metric properties (e.g., degree of curvature, length or aspect ratio) can be ignored. Consistent with this interpretation, the metric variation that was produced by rotation in these experiments had very slight effects on recognition performance.

The invariance obtained in the experiments with unfamiliar stimuli (all the experiments) and unfamiliar object classes (3, 4, and 5) suggest that the large effects of rotation angle on RTs found in previous work with unfamiliar stimuli (e.g., Edelman & Bülthoff, 1992; Tarr, 1989) was a consequence of the use of stimulus sets whose members failed to satisfy the conditions for invariance rather than being the result of unfamiliarity per se. The absence of an effect of rotation in depth despite large changes in global shape (particularly in Experiment 1) provides strong evidence against the position argued by Cave and Kosslyn (1993) in which a central role was assigned to global shape in object recognition.

A major result from this investigation is that whether a given rotation will cause a large decrement in performance depends on whether the rotation produces a change in the GSD for that object. Some investigators have interpreted an average monotonic (deleterious) effect of orientation disparity as evidence for the build up of viewpoint-specific templates (e.g., Edelman & Bülthoff, 1992; Rock & DiVita, 1987; Srinivas, 1993; Tarr, 1989). As noted in the Discussion of Experiment 2, unsystematic selection of stimuli is likely to produce just such a function as geons are occluded or revealed at the different orientations. Where template theories have been well specified, as that of the elastic templates of Poggio and Girosi (1990) and Poggio and Edelman (1990), the theoretical entities specify individual objects at specific orientations. A much more critical test of such theories, therefore, would involve the correlation of the magnitude of the rotation-disparity decrement with the distances in the representation space for individual objects at the particular studied and tested orientations rather than the average effect of rotation angle.

The Presence of Small Rotation Effects

Although the effects of orientation disparity on RTs or error rates were extremely small in those conditions in which the stimuli could activate distinctive GSDs, none of the experiments (other than the single geon study) resulted in a complete absence of an effect of orientation disparity, and in one (Experiment 4) the increase in error rates with increasing rotation angle was significant. (Perhaps counteracting that effect was the reliable decrease in error rates with increasing rotation angles in Experiment 3.) How should these effects be regarded? One possibility is that there are alternative ways of classifying a shape and though the account provided by geon theory might characterize much of the processing, on some trials subjects may select some other basis to organize their responses. In fact, if one can instruct a person to perform a given classification task on the basis of a particular kind of information, then that information might be used at any time. In particular, some subjects on some trials might have opted to respond on the basis of a viewpoint-specific aspect of the stimulus, such as global shape. For example, it would be possible to classify a stimulus on the basis of whether it is elongated. Similarly, subjects (sometimes) might have used a viewer-centered representation as to the location of a particular object part or feature, a mode of processing that might be more characteristic of dorsal system functioning. In all of these cases, at zero rotation disparity, the viewpoint-specific information would be valid and correct responses would be facilitated. The negligible effects of rotation disparity (where the invariance conditions were met) suggest that such modes of processing were only rarely selected, if at all. Last, it is possible that the effects of rotation could have been merely a consequence of uncontrolled foreshortening or occlusion, which would have the effect of reducing the activation of the original GSD. An important research question, from the theoretical perspective advanced here, is the need for a principled quantitative analysis to determine the costs in activation level (and latency) when a GSD is changed. Such an analysis would have to include a resolution function in that a part need not completely appear or disappear as a result of an orientation change before that change will begin to affect performance. Undoubtedly, the modeling of costs would have to include the effects of the similarity of neighboring objects.

A related but perhaps more difficult problem is how to conceptualize contour that is near the category boundary of a nonaccidental contrast, such as an edge that is very slightly curved. Edelman (1993) recently argued for a continuum of shape similarity, in which more similar shape classes (on the basis of transduction of receptive fields) show greater dependence on viewpoint, with no particular status of nonaccidental contrasts. Although greater similarity generally would be expected to produce steeper effects of rotation from almost any theoretical account, there is a benefit of nonaccidental differences between categories that greatly exceeds what would be expected from the direct matching of the outputs of receptive fields (Cooper & Biederman, 1993; Fiser, Biederman, & Cooper, 1993).

Implications

The central issue regarding viewpoint invariance in three-dimensional object recognition is the prediction made by different theories concerning the behavior of the system when one is required to identify an object following a change in depth orientation. Geon theory predicts that viewpoint invariance can be achieved for all views that activate the same geon structural description. Alternatively, viewpoint-dependent theories, particularly those assuming point-to-point matching of templates (such as those proposed by Edelman & Bülthoff, 1992; Lowe, 1987; Tarr, 1989; and Ullman, 1989) assume that for unfamiliar objects the appearance of invariance is achieved by a mechanism (such as mental rotation, extrapolation, interpolation, or alignment) sensitive to rotation disparity, which suggests that the time needed to identify an image will increase monotonically as the viewed orientation diverges from a learned orientation.[12] In the case of familiar objects, according to the viewpoint-dependent position, invariance is achieved by a multiplicity of stored views. The five experiments presented here provide evidence for unfamiliar representations that can be recognized largely independently of changes in depth orientation.

[12] A special case arises with Ullman's (1989) alignment model. This model can readily achieve viewpoint invariance in that it does not require that recognition times be proportional to disparity between learned and probed views (though Ullman has cited mental rotation costs as evidence for an alignment process). If no performance cost is assigned to rotation angle in depth, then in many cases where there is a cost for rotation in depth with plenty of alignment points present (such as the parts change condition of Experiment 3), one would not be produced. Moreover, the alignment scheme need not produce any costs for rotation in the plane or nonaffine transformations such as reflection. Unlike human subjects, the model would readily demonstrate immediate viewpoint invariance for the original Edelman et al. (1993) bent paper clip stimuli and, like the other template models cited, would find it especially difficult to recognize a novel object on its very first presentation.

These results add to previous findings of invariance in name priming over mirror reflection (Biederman & Cooper, 1991a, 1991b; Cooper, et al., 1992) to challenge the claim that recognition is achieved through viewpoint-specific instances.

How should we interpret the relatively shallow slopes over rotation angle after extensive practice in the Edelman and Bülthoff (1992) experiments on classifying bent paper clip objects (and others of that type)? As noted previously, it is possible that what the subjects were learning while performing their task are (what normally would be) nonaccidental descriptions of the stimuli, but because of the accidents, these descriptions are distinctive for only a small range of orientations. That is, the stimuli produced different GSDs with slightly different views, thus failing to satisfy Condition 3. The various descriptions for a given object could then have been linked, perhaps in a nonvisual associative memory, much as one might learn what the front and the back of a particular house might look like. If any of the configurations suggested a familiar object, that code could have been learned as well. Consider the three images shown in Figure 4 depicting three poses of an object like those used in Edelman et al. (1989). Pose A might be coded as a tilted dreidel (Hanukkah top), Pose B as a tilted glass with a bent straw, and Pose C as a headless man sitting on the ground with his arms extending over his bent knees. Much of the learning in this task might have occurred to determine which characterizations went with which objects; that is, that the dreidel, glass, and man were all object No. 7. From this account, for all three images, GSDs would be activated that would normally allow viewpoint-invariant recognition but—because of the high likelihood for wire-frame objects to produce violations of Condition 3—invariance was not obtained. A representation specifying the exact angles and lengths of the stimuli, as posited by Edelman and Bülthoff (1992), need not be invoked.

However, with other stimuli, as with the ameboid blobs of Edelman and Bülthoff (1992), metric variation is, perhaps, all that is available for recognition. Performance with such stimuli early in practice is, predictably, atrocious. Edelman and Bülthoff (1992) argued that their tasks with metrically varying stimuli reflect those processes that are used when making subordinate, rather than basic-level, categorizations. If this is true, it may be true for only a tiny proportion of the subordinate-level classifications that people make. Most of the time people have little difficulty distinguishing among different models of chairs, for example. When one is faced with the task of distinguishing among highly similar exemplars in difficult subordinate-level classification tasks, one typically achieves high levels of accuracy, not by creating precise templates, but by discovering (or asking) where to look and what viewpoint-invariant contrast to seek (Biederman & Shiffrar, 1987). For example, the discriminating information given in bird books is virtually never specified metrically but is typically specified as a viewpoint-invariant shape or surface (color or texture) feature or a small set of features, at a fine scale. The presentation of the distinguishing information is contingent on first resolving the entry level of the bird, for example, that it is a duck. Similarly, when one tries to determine if a given car is a Mazda 626, a Honda Accord, or a Toyota Camry, one looks for the symbol (or name) that the logo designers conveniently distinguish from other logos on the basis of nonaccidental contrasts.

We know that this type of learning, which can be described as a discovery of the locus of small diagnostic nonaccidental contrasts, occurs (Biederman & Shiffrar, 1987). Future research will determine whether such learning is sufficient for all cases of rapid subordinate classification. For now, we conclude that when objects activate distinctive GSDs, the speed and accuracy of their recognition suffers little—if at all—from rotation in depth.

References

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

Beck, J. (1967). Perceptual grouping produced by line figures. *Perception & Psychophysics, 2*, 491–495.

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

Biederman, I., & Cooper, E. E. (1991a). Evidence for complete translational and reflectional invariance in visual object priming. *Perception, 20*, 585–593.

Biederman, I., & Cooper, E. E. (1991b). Priming contour-deleted images: Evidence for intermediate representations in visual object priming. *Cognitive Psychology, 23*, 393–419.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 121–133.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 640–645.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a 2-D view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, 89*, 60–64.

Cave, C. B., & Kosslyn, S. M. (1993). The role of parts and spatial relations in object identification. *Perception, 22*, 229–248.

Cooper, E. E., & Biederman, I. (1993, May). *Metric versus viewpoint invariant shape differences in visual object recognition.* Poster presented at the annual meeting of the Association for Research in Vision and Ophthalmology, Sarasota, FL.

Cooper, E. E., Biederman, I., & Hummel, J. E. (1992). Metric invariance in object recognition: A review and further evidence. *Canadian Journal of Psychology, 46*, 191–214.

Dickinson, S. J., Pentland, A. P., & Rosenfeld, A. (1992). From volumes to views: An approach to 3-D object recognition. *Computer Vision, Graphics, and Image Processing: Image Understanding, 55*, 130–154.

Edelman, S. (1993). *Class similarity and viewpoint invariance in the recognition of 3D objects* (Tech. Report CS92-17). Rehovot, Israel: Weizmann Institute.

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of 3D objects. *Vision Research, 32*, 2385–2400.

Edelman, S., Bülthoff, H., & Weinshall, D. (1989). *Stimulus familiarity determines recognition strategy for novel 3D objects* (Artificial Intelligence Laboratory Technical Report No. 1138). Cambridge, MA: MIT Press.

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

Eggert, D., & Bowyer, K. (1990). Computing the orthographic projection aspect graph for solids of revolution. *Pattern Recognition Letters, 11,* 751–763.

Ellis, R., & Allport, D. A. (1986). Multiple levels of representation for visual objects: A behavioural study. In A. G. Cohn & J. R. Thomas (Eds.), *Artificial intelligence and its applications* (pp. 245–247). New York: Wiley.

Fiser, J., Biederman, I., & Cooper, E. E. (1993). *To what extent can matching algorithms based on direct outputs of spatial filters account for human shape recognition?* Unpublished manuscript, University of Southern California, Los Angeles.

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

Humphrey, G. K., & Kahn, S. C. (1992). Recognizing novel-views of three-dimensional objects. *Canadian Journal of Psychology, 46,* 170–190.

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

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

Koenderink, J. J. (1990). *Solid shape.* Cambridge, MA: MIT Press.

Kohlmeyer, S. W. (1992). Picture perception lab: A program for picture perception experiments on the Macintosh II. *Behavior Research Methods, Instruments, & Computers, 24,* 67–71.

Kosslyn, S. M. (in press). *Image and brain: The resolution of the imagery debate.* Cambridge, MA: MIT Press.

Kosslyn, S. M., & Koenig, O. (1992). *Wet mind: The new cognitive neuroscience.* New York: Free Press.

Kriegman, D., & Ponce, J. (1990). Computing exact aspect graphs of curved objects: Solids of revolution. *International Journal of Computer Vision, 5,* 119–135.

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

Palmer, Rosch, & Chase. (1981). Canonical perspective and the perception of objects. In J. Long and A. Baddeley (Eds.), *Attention and performance* (vol. 9, pp. 135–151). Hillsdale, NJ: Erlbaum.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize 3D objects. *Nature, 343,* 263–266.

Poggio, T., & Girosi, F. (1990). Networks for approximation and learning. *Proceedings of the Institute of Electrical and Electronic Engineers, 78,* 1481–1497.

Poggio, T., & Vetter, T. (1992, February). *Recognition and structure from one 2D model view: Observations on prototypes, object classes and symmetries* (MIT Artificial Intelligence Laboratory Tech. Report 1347).

Rock, I. (1973). *Orientation and form.* San Diego, CA: Academic Press.

Rock, I., & DiVita, J. (1987). A case of viewer-centered perception. *Cognitive Psychology, 19,* 280–293.

Rock, I., Wheeler, D., & Tudor, L. (1989). Can we imagine how objects look from other viewpoints? *Cognitive Psychology, 21,* 185–210.

Shepard, R. N., & Cooper, L. A. (1983). *Mental images and their transformations.* Cambridge, MA: MIT Press.

Srinivas, K. (1993). Perceptual specificity in nonverbal priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 582–602.

Tarr, M. J. (1989). *Orientation dependence in three-dimensional object recognition.* Unpublished doctoral dissertation. Massachusetts Institute of Technology, Cambridge, MA.

Treisman, A. M. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology, 40A,* 201–237.

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

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

Vetter, T., Poggio, T., & Bülthoff, H. H. (1993, May). *Recognition of symmetric 3D objects.* Poster presented at the annual meeting of the Association for Research in Vision and Ophthalmology, Sarasota, FL.

Received April 6, 1992
Revision received June 11, 1993
Accepted January 8, 1993 ■

Correction to Ballas (1993)

The article "Common Factors in the Identification of an Assortment of Brief Everyday Sounds," by James A. Ballas (*Journal of Experimental Psychology: Human Perception and Performance,* 1993, Vol. 19, pp. 250–267), is in the public domain. A previous notice regarding this article ("Correction to Ballas (1992)," *Journal of Experimental Psychology: Human Perception and Performance,* 1993, Vol. 19, p. 829) incorrectly identified the volume number and date of publication of the journal in which this article appeared.