# Recognizing Depth-Rotated Objects: A Review of Recent Research and Theory

## Irving Biederman

Department of Psychology and Neuroscience Program
Hedco Neurosciences Building, MC2520, University of Southern California
Los Angeles CA 90089-2520 USA

## Abstract

Many of the phenomena of object classification can be derived from a representation specifying a nonaccidental characterization of an object's parts (geons) and relations, termed a geon structural description (GSD). Such a representation: a) enables the facile recognition of depth-rotated objects, even when they are novel, b) provides the information that is employed not only to distinguish basic-level but also highly similar members of subordinate-level classes, and c) enables mapping onto verbal and object-reasoning structures. Recent psychophysical and neural investigations of object recognition have provided additional support to this theory of object representation.

## I. Introduction

Theories of the representation mediating object recognition are often termed "controversial," particularly in accounting for the effects of rotation in depth. The apparent controversy centers on whether "view-based" templates (e.g., Poggio & Edelman, 1990) are to be preferred to structural descriptions that posit simple viewpoint-invariant part primitives (geons) and explicit categorized relations (i.e., geon structural descriptions, GSDs) (Biederman, 1987, Hummel & Biederman, 1992). GSDs place heavy reliance on nonaccidental properties (NAPs), which are largely unaffected by rotation in depth. NAPs provide a ready basis for distinguishing one object's parts and relations from another. In contrast, neither NAPs nor explicit parts nor explicit relations are specified by view-based templates. We note here that some of the leading proponents of view-based templates (e.g., Riesenhuber & Poggio, 1999) have very recently abandoned templates in favor of feature lists in which features are translationally and scale invariant. This development is discussed in a later section.

After an initial demonstration contrasting object recognition with and without distinctive GSDs, recent empirical and theoretical work on the issue of the recognition of depth-rotated objects will be considered. Insofar as the evidence for the employment of NAPs (and parts) provided a strong evidential basis for the abandonment of templates, this review is still timely.

### 1.1 A Demonstration

Imagine performing a matching task in which you are to determine if two sequentially presented novel objects are the same or different, irrespective of their orientation in depth. Before scrutinizing Figure 1, please cover the objects with your hand. The figure illustrates some possible trials in which the object on the left is always S1, the first object. The objects in the right column are possible S2s. Take a quick peek at S1. You probably can describe it. Now take a quick peek at the top object in the S2 column. It should be trivially easy to respond "different." The same would be true of the second object in the S2 column. Or the third. You might judge the bottom object to be "same," even though the object is now rotated in depth, as is

the wedge in the previous object.  The first three trials differed in at least one geon and the discrimination is trivially easy.  The last object had the same geons and it looks the same as S1, despite its depth rotation.  Little or no rotation costs would be expected with such objects.

**Fig. 1**.  (Next page.)  An illustration of four trials in a Same-Different matching task of two-geon novel objects taken from Biederman and Bar (1999).  The object on the left is S1, the first stimulus for all four trials.  The four objects in the right column are possible S2s.  The top object differs in both geons; the second and third in one geon, and the bottom object is the same, but rotated in depth.  Observers should have no trouble accurately performing same-different judgments.  Nor should they have any difficulty in describing the objects and how they differ from each other. Only the third and fourth S2s would have been trials in the Biederman and Bar (1999) experiment.
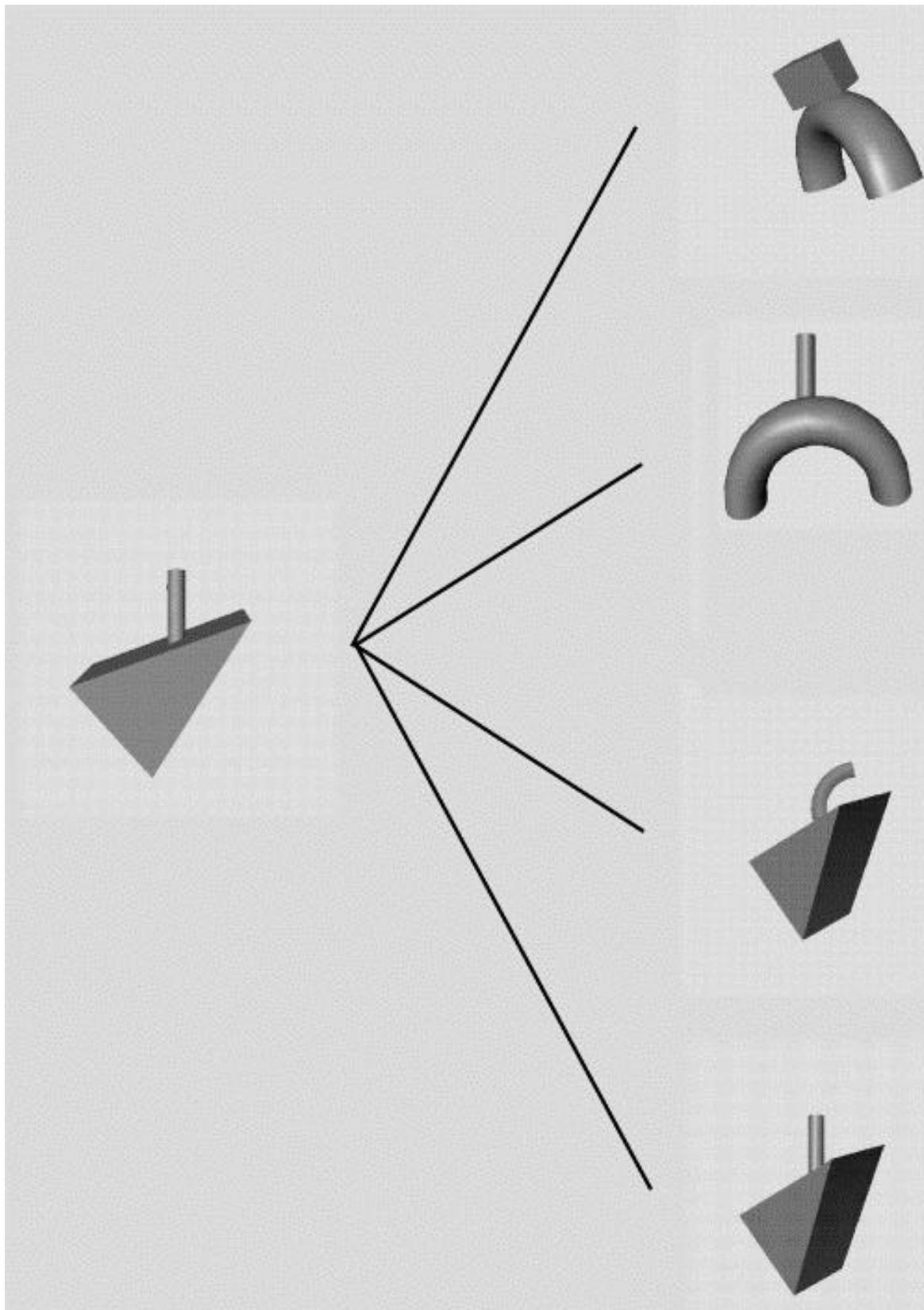
In contrast to the types of objects shown in Figure 1 are the bent paper clips (and stimuli with similar characteristics) where neither the parts nor the relations (the angles between the segments for such objects) are distinguished by NAPs. NAPs are to be distinguished from metric properties (MPs), such as aspect ratio of a part or the angle between parts, which do vary with orientation in depth. Figure 2 allows the reader to try to match such stimuli, used extensively in the study of object recognition under variations in orientation, under the same conditions as used for the objects in Fig. 1. If you are like most people, you experienced considerably more difficulty in performing this task. Many judge the top three S2s to be 'same' and the bottom one to be 'different'. This is precisely wrong. Not surprising, enormous rotation costs are incurred in matching such objects (e.g., Edelman & Bülthoff, 1992).

Given the ease of matching the objects in Figure 1 compared to the extraordinary difficulty in matching those in Figure 2, it is a wonder that there is any controversy at all about the representations mediating object recognition over depth orientations.

But what is the controversy? All theories are view-based. Whereas GSDs are a theory of the representation of an object, "view-based" is merely an effect—a cost of rotation--that distinguishes no current theoretical alternative. That there would be costs of rotation was explicitly discussed by Biederman (1987) who noted that rotation in depth would alter the GSDs, as parts were occluded or revealed, and that a cost function was required based on the similarity of the GSDs. Insofar as I also noted that different representations (i.e., different GSDs) would be required for substantially different views, I probably should be regarded as the originator of view-based theories.

To the extent that "view-based" is taken to be a specific template representation of the kind proposed by Poggio & Edelman (1990); Edelman, (1995) and Ullman (1996 to account for costs in the recognition of a set of depth-rotated bent paper clip stimuli, they can be evaluated against GSDs. It must be acknowledged that neither competing account of object recognition is complete. For example, the Poggio & Edelman model (also employed by Logothetis, Pauls, Bülthoff, & Poggio, 1994) by which vertices of the clip segments are mapped onto a set of radial-basis classifiers does not specify how the classifiers can be ordered to correspond to the specific vertices. If the segments were connected to the same vertices in a different order than a vastly different appearing clip would be defined.

**Fig. 2**. (Next page.) An illustration of possible trials in Biederman and Bar's (1998) Same-Different matching task with bent paper clips. The scheme is the same as in Fig. 1. Contrary to initial impressions, the top three S2 objects are different from S1; the bottom one is the same. In Biederman and Bar's (1998) experiment, high false alarm rates were associated with the top three S2s and high miss rates with the bottom S2.
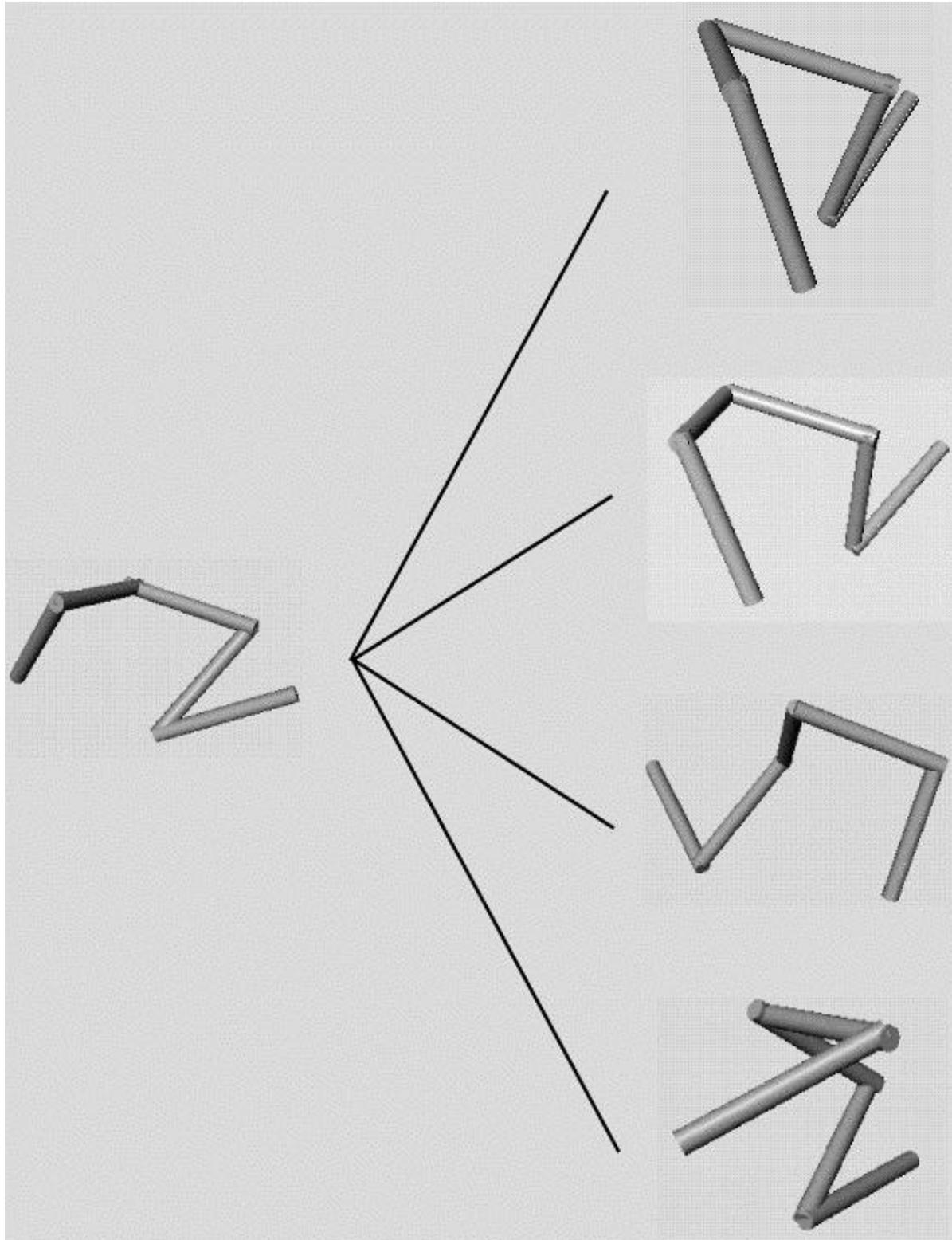
Biederman and Hummel (1992) provide a number of images where their implementation of geon theory, JIM, would fail to account adequately for the perceived object. Computer vision engineers are apt to criticize generalized cylinder representations (of which geons are a special case) on the grounds that it is difficult to extract generalized cylinder models of objects from gray level images.  Indeed, JIM's input was limited to line drawings (and abstracted ones at that). However, such criticism may be more appropriately directed to the state of the art of current edge finders than to biological representations.  People have no trouble determining the orientation and depth discontinuities—the important edges for GSDs--of novel objects and scenes.  They readily distinguish such edges from those produced by reflectance variations, shadows, and surface markings.  So we have an existence proof that rapid, accurate bottom-up edge extraction is possible and done billions of times a second by human brains on this planet. Recently Zerroug and Nevatia (1996) have shown that the generalized cylinders comprising a complex object can be extracted from a single gray-level image.

## 2.  Current Accounts of Visual Object Classification

Another shortcoming of GSDs is that they require a solution to the binding problem.  A concern is that the binding mechanism of JIM—correlated firing—may to be too slow to produce the 100 msec real-time object and scene recognition that is so evident in human performance. JIM activated geon representations from the synchrony of firing of the features that comprised each geon.  Hummel and Stankiewicz (1996) perhaps offer some progress along this front with a hybrid model that assumes spatially enumerated templates of the geons.

Although a template representation, of the kind proposed by Lades, Vortbrüggen, Buhmann, Lange, von der Malsburg, Würtz, & Konen (1993), may well characterize the earlier stages of cortical processing, from V1 through V4, the later stages would appear to map more readily into GSDs, as discussed below.  By making part types and relations explicit, GSDs allow ready mapping of such descriptions to language and structures that can mediate reasoning. When the reader first looked at S1 in Fig. 1, no doubt he or she could easily describe the object as something like "a vertical cylinder on a tilted wedge."  (In general, people describe objects in terms of their parts, Tversky & Hemenway, 1984.)  Note that the relation, "on," was explicit in the description.  One can also readily characterize, and hence reason about, the differences between S1 and the first three S2 objects, e.g., the third S2 object differs from S1 in that its top part (i.e., what is "on") is a curved rather than a straight cylinder.  It is not at all clear how to do such mapping with templates or feature lists-- a serious shortcoming of such accounts as characterizations of human object cognition.  Whereas metric templates may be inadequate for object representation, they likely are the representation mediating face recognition (Biederman & Kalocsai, 1997; Wiscott, Fellous, Krüger, & von der Malsburg, 1997).  With respect to the current discussion, Biederman and Kalocsai (1997) note how difficult it is for people to describe the differences among a set of similar faces that they can otherwise readily discriminate.

## 3.  NAPs (geons) vs. MPs:  Experimental Results

GSDs specify both parts and relations.  I will here concentrate on the NAP characterization of the parts (see Biederman, 1995, for a summary of the evidence supporting the role of simple parts in the representation mediating visual object recognition).

Distinctive NAPs can confer an enormous benefit in attempting to determine whether two bent paper clips are the same or different when they are viewed at different orientations in depth (Biederman & Gerhardstein, 1993). These investigators substituted a different geon for each center segment of a set of 10 line drawings of bent paper clips.  The addition of the distinctive geon dramatically reduced rotation costs (to 5,000°/sec) from a level with error rates so high that RTs were, essentially, uninterpretable.

View-based template accounts, in assigning no special status to NAPs or parts, would require familiarity with the specific views of novel objects, with only a modest generalization gradient around a nearby view.  Some (Tarr & Bülthoff, 1995) protested this demonstration, arguing that NAPs were of value only with a small set of known stimuli where people could anticipate a distinguishing NAP.  That people would spontaneously exploit distinguishing NAPs was, indeed, one of the points that Biederman and Gerhardstein wished to make, but is familiarity required to get immediate viewpoint-invariance with novel objects?

Moshe Bar and I (Biederman & Bar, 1999) compared directly the rotation costs for detecting differences in either a Metric Property (MP) or a NAP in a same-different sequential matching task.  We used novel, rendered two-part objects, such as those shown in Fig. 1, presented at either the same or different orientations-in-depth.  On half the trials the objects were identical; on half the trials they differed in either an MP, e.g., aspect ratio, of a single part or a NAP, e.g., straight vs. curved axis (producing a different geon) of a single part.  The contrast of the object on the left and the third object in the right hand column of Fig. 1 illustrates a NAP difference (straight- vs. curved-axis cylinder).  The MP variation would have been a cylinder with a different length (aspect ratio) or angle of attachment to the wedge. The subjects saw a given stimulus sequence only once, so they could not predict whether a part would change, or, if there was a change, which part would change and how it would be changed.

How much of an MP difference is equivalent to a given NAP difference?  To address this apples-and-oranges problem, the MP and NAP differences were selected to be equally discriminable (as assessed by RTs and error rates) when the objects were at the same orientation in depth (0° orientation difference).  The MP image differences were also approximately the same magnitude (actually slightly larger) then the NAP image differences when the images were scaled according to a wavelet-like similarity measure (Lades, 1993).  Rotation angles that averaged 57° produced only a 2% increase in error rates in detecting the NAP differences but a 44% increase in error rates (to a level that was below chance) in detecting MP differences.

The benefit conferred by NAPs, documented by Biederman and Gerhardstein (1993), and confirmed by Biederman and Bar (1999) is quite dramatic and is among the largest effects in shape recognition.   Another enormous effect is the difference in recognizability between recoverable and nonrecoverable contour deletion (Biederman, 1987).  Neither of these effects are handled by metric template accounts.

Nonetheless, rotation costs, though small, are often apparent even with distinguishing GSDs (Biederman & Gerhardstein, 1993; Biederman & Bar, 1999; Hayward & Tarr, 1997; Tarr, Bülthoff, Zabinski, & Blanz, 1997; Tarr, Williams, Hayward, & Gauthier, 1998).  What might be producing these costs?  It is possible, as noted by Biederman and Gerhardstein (1993), that an orientation-specific representation underlies these costs.  This representation may be of one of two types, given current theorizing:  a) an episodic representation that binds view information along to an invariant representation of shape, as detailed by Biederman and Cooper (1992, see below), that could be employed on some percent of the trials to mediate performance (though not necessarily object perception), or b) that there are viewpoint specific representations directly mediating object perception.  But before the latter alternative is accepted merely on the basis of some costs with distinguishing GSDs present, other possible bases for the costs must be ruled out.

Look again at Fig. 1 and consider what one would have to do to make it difficult, under rotation, to determine that the third S2 object was different from the first and the fourth S2 object was the same.  One way would be to render the object in such a way that it would be difficult to determine if the distinguishing geon was curved or straight.  Biederman and Bar (1999), in a

critical review of those studies reporting high rotation costs, noted that low resolution of the distinctive geon was a common characteristic in such studies. Biederman and Bar (1998) showed that factors that increased the discriminability of distinguishing geons in rendered images, such as avoiding near accidents or using increased exposure durations, greatly reduced rotation costs

There is another, more subtle, effect that could have contributed to the apparent costs of rotation in the Tarr et al. (1997; 1998) and Hayward and Tarr (1997) same-different matching studies.  On rotated Same trials and all Different trials in a Same-Different matching task, a "difference" signal might be produced by the change in luminance of specific display positions. This signal may be related to Nowak and Bullier's (1997) finding that marked changes—a transient--in a stimulus produce a signal that rapidly propagates through the ventral pathway all the way to frontal cortex.  (Because of the intervening mask, the difference signal would be between S2 and an actively maintained representation of S1, as noted by Biederman and Bar, 1999.)  No difference signal would be produced when S1 and S2 are the same, unrotated, object in the same position.  So the subject could readily use the *absence* of a difference signal to respond Same on unrotated (0°) trials, artifactually lowering reaction times (RTs) on such trials with the consequence that the slope of the RT X Rotation Angle function would be increased. Biederman and Bar (1998) showed that the effect of this artifact in increasing rotation costs could be greatly reduced by merely shifting S2 with respect to S1 on all trials, so that the difference signal was always present.  The translation, by producing a difference signal on all trials, served to raise the RTs and error rates for 0° trials relative to rotated trials.  This had the effect of greatly reducing the *apparent* costs of rotation.  Biederman and Bar's (1999) experiment, which found near invariance over rotation, also translated S2 with respect to S1 on all trials.

Despite the length of the preceding paragraphs as to the origin of rotation costs when distinctive GSDs are available, it must be emphasized that these rotation costs (against zero costs) are small relative to the gigantic costs that are evident when distinctive GSDs are not present.  We also note that even view-based theorists always find that objects with distinctive geons show lower rotation costs than those without distinctive geons (Tarr et al., 1997; Logothetis, et al., 1994).  Tarr and his associates had suggested that "mental rotation" is required to match such stimuli although this position has recently been abandoned by Tarr (Hayward & Tarr, 2000).  It is difficult to believe that mental rotation (whatever its ultimate nature), with its attendant high costs, would be required to judge the objects in figure 1.

*3.1 Can view-based accounts incorporate geons as a unique or diagnostic feature?*

A reviewer, responding to my earlier point that "view-based accounts assign no special status to NAPs," asked whether view-based theorists could regard geons as some kind of unique or diagnostic feature extracted from a 2D view.  The answer is, "of course."  But there is a serious problem with an account that holds that a unique or diagnostic feature will be employed for recognition.  How does the perceiver know what is unique or diagnostic the first time he or she views an object?  Consider, again, an individual seeing the nonsense object on the left side of Fig. 1.  The coding of that object would, roughly, appear to be a vertical cylinder on top of a wedge.  That is, the object is described in terms of its simple parts and the relations among these parts (Tversky & Hemenway, 1984).  This type of representation, a geon structural description (GSD), may well be the default description that the visual system generates in the absence of explicit knowledge about the other to-be-discriminated objects.  GSDs often convey the functionality—or affordances—of the object.  Moreover, GSDs often readily map on to verbal descriptions and allow reasoning about objects.  We can readily say how the four objects on the right side differ from the one on the left (or from each other).

The important question is not whether a representation is view-based, but what that representation is (as, again, *all* representations are view-based). The phenomena of: a) small rotation costs with distinctive GSDs b) the sizable costs in recognizing new views of objects, such as a set of bent paper clips, that are not distinguished by GSDs (as discussed in the next section), and c) the reduction in the costs in b) from learning the new views, has obscured the issue of representation, insofar as the nature of what was learned was often not considered. In allowing translational and scale feature invariance, the recent Riesenhuber and Poggio (1999) scheme resembles an earlier proposal by Bartlett Mel (19997). There is nothing in the Riesenhuber and Poggio model to suggest the enormous inferential leverage and invariance to rotation costs provided by distinctive NAPs or the difference in recognizability between recoverable and nonrecoverable contour deletion. These models are, essentially, feature lists in that they do not posit explicit structures, such as parts and relations among parts, by which objects might be represented--and described. Instead, different arrangements of the parts merely produce new features. A potential serious shortcoming of such models is that it is not clear how well they would do with novel objects that are to be distinguished from unknown sets of other objects, such as with the task illustrated in Figure 1.

### 3.2  Observations about bent paper clips as experimental stimuli

The central motivation for devising bent paper clip stimuli (and others of similar design) was that they would be unfamiliar, so that the learning of different poses could be studied. However, one must consider an obvious characteristic that accounts for much of the extraordinary difficulty in classifying members of sets of such stimuli: They lack of distinguishing GSDs.

The absence of distinguishing GSDs in the standard set of bent paper clips means that the critical information for everyday shape recognition is absent from these stimuli so the relevance of such objects to normal recognition can be questioned. Some bent paper clip devotees have suggested that their stimuli are relevant for subordinate-level recognition, such as the difference between different kinds of tables. However, a review of the vast majority of subordinate-level classifications that people make in their lives suggests that it is extremely rare that distinguishing GSDs are not available. A square table can be distinguished from a round table without appeal to metric information and certainly without engaging in mental rotation. Biederman, Subramaniam, Bar, Kalocsai, and Fiser (1999) note that NAPs of small regions, rather than metric templates, are specified for discriminating among highly similar classes such as birds on the same page in the bird guides.

Think of how you would discriminate two different chairs of the same manufacturer's model. Without fail, visitors to my office look for a distinctive scratch or stain or other such *nonaccidental* difference, at a small scale. They never consider what is readily expressed by metric templates--a template of the whole chair or, in selecting a small feature, those that might differ metrically (at a modest scale).

The objects shown in Fig. 1 meet the criteria of being unfamiliar, yet in retaining distinctive geons they allow one to study how such information might be used. Although a set of paper clips lack distinctive GSDs, their projections often provide an accidental or near accidental characteristic that people try to interpret in terms of GSDs (Biederman & Gerhardstein, 1993; 1995). For example, the bottom S2 object in Fig. 2 resembles an arrow that would normally be produced by actual cotermination of segments but is here an accident of viewpoint. Biederman and Bar (1998) observed that when there were such qualitative differences in appearance—typically well captured by differences in GSDs—miss rates were extremely high. When S1 and S2 were actually different clips but with similar GSDs, as in the upper three S2s of Fig 2, then the false alarm rates were extremely high.

As rotation angles increase from 0 to 90°, there is an increasing chance of changes in the qualitative characterization of such stimuli.  The oft reported increase in matching costs with increasing rotation angles may be more a consequence of an increasing chance for a change in an accidental GSD then in the rotation of a template.  Consistent with this interpretation are the low rotation costs for 180° rotations.  Such rotations often approximate mirror reflections under which the GSDs are preserved.

*3.3 When GSDs are Insufficient*

There is no doubt that aspects of early cortical representation are well described by a 2D array of local filters at a variety of scales and orientations.  The view expressed here is that the outputs of such a representation are mapped onto nonaccidental classifiers—such as units distinguishing straight from curved lines or various vertices produced by cotermination of end stopped activity.  A vector representing the activity of these nonaccidental classifiers (which, in JIM, are bound through correlated firing), in tern, activate units akin to Hummel and Biederman's (1992) geon feature assemblies, representing single or pairwise combinations of geons and their invariant nonaccidental relations, such as VERTICAL_CYLINDER_ABOVE_PERPENDICULAR_SMALLER_THAN_X.  The output of such geon feature assemblies could readily map onto language, as evidenced by the manner in which people describe the objects in figure 1, as well as memory structures supporting object cognition.

What if the stimulus does not have distinguishing parts and nonaccidental properties, as with the set of smooth blobby shapes studied by Shepard and Cermak (1973)?  In such a case the nonaccidental classifiers would not be differentially activated to distinguish the members of the stimulus set and the observer would have to rely on whatever metric information distinguished the stimuli, in which case the similarity space would be that established by the early local, multiscale, multioriented Gabor-like filters (Biederman & Subramaniam, 1997).  It should also be the case that discrimination among such stimuli should be more difficult than if distinctive GSDs were available (at the same level of spatial filter similarity), show more rotation costs, be difficult to articulate, and not be the basis of natural concept distinctions.

Discrimination performance among a set of highly similar faces shows such characteristics (Biederman & Kalocsai, 1997), as well as similar pairs of the Shepard and Cermak (1973) shapes (Biederman & Subramaniam, 1997) and objects with irregular parts (Cooper, Subramaniam, & Biederman, 1995).  See Biederman (1995) for a review.

## 4.  Recent Neural Evidence for GSDs

It has long been known that macaque inferior temporal (IT) neurons are highly shape selective and that different neurons show different shape preferences.  Tanaka (1993) demonstrated that these preferences can be elicited quite strongly to features of  "moderate complexity," typically composed of one or two parts.   This level of complexity closely matches what would be expected from single geons, invariant shape features, and, most frequently, geon feature assemblies (Hummel & Biederman, 1992), in which two geons are bound in a specific relation.   To a first approximation, Tanaka's (1993) and Kobatake and Tanaka's (1994), moderately complex features would seem to be viewpoint invariant.  Consistent with this interpretation is Esteky and Tanaka's (1998) results showing that metric variation, viz., changes in aspect ratio that would be produced by a rotation in depth, had only a minimal effect on IT cell activity.

Vogels, Biederman, Bar, and Lorincz  (In press) tested macaque IT (area TE) neurons with the identical set of two-geon stimuli used by Biederman and Bar (1999) to determine if greater modulation in cell activity would be produced by a change in a geon compared to a change in an MP (Metric Property). They found that geon changes, despite their smaller image changes (as assessed by wavelet similarity measures), produced greater modulation (up or down) in cell activity.  Moreover, when the original objects were rotated (i.e., those without an MP or geon change), the modulation attributable to the rotation itself was highly correlated with the modulation produced by MP changes for that cell but completely independent of the modulation produced by the geon changes.  Such a tuning pattern would be expected given the results of Biederman and Bar (1999) that only geon-changed stimuli were readily discriminable from the originals under rotation.

There have been a number of reports of TE cell preferences reflecting experimental manipulations of familiarity (e.g., Logothetis, et al., 1994; Tanaka, 1996).  There are two points to be made about such demonstrations.  First, tens, if not hundreds, of thousands of trials are required to obtain such preferences (Logothetis, 1999).  Second, as discussed previously, it is not unlikely that there are at least two functions of object recognition subserved by IT.  One is to provide descriptions of objects, novel or familiar, such as what the reader experienced when first viewing S1 in Fig. 1.  Such a system is likely well developed by late infancy.  The second function is to provide an episodic record of the perceptual experience with particular objects or scenes. It is possible that the cells found in the training experiments are those subserving this second episodic memory function.  That there may be these two representations of objects was documented by Biederman and Cooper (1992) who showed that the priming of object naming was invariant with size changes but that such changes produced considerable interference on episodic old-new judgments of the shape of the object (in which size was to be ignored). Distractors in that experiment were objects with the same name but a different shape.  Similar results were found for changes in position and reflection (Biederman & Cooper, 1991) and orientation (Cooper, Biederman, & Hummel, 1993).  Although the first function probably supports lion's share of human object recognition, it would certainly be possible to employ the second problem to solve particular classification tasks.  If I know that a chair is on the right and a table on the left, a flash of an object on the right could lead me to infer that it was a chair rather than a table.  Such view information could be employed whenever there was difficulty in determining an object's GSD.

## 5.  Conclusions

The evidence suggests that GSDs provide a suitable representation with which to understand the large body of results that have recently accumulated in the study of depth rotated objects.  In addition, GSDs provide a basis for understanding the general problem of object perception and reasoning.

### References

Biederman, I.  (1995).  Visual object recognition.  In S. M. Kosslyn and D. N. Osherson (Eds.). *An Invitation to Cognitive Science*, 2nd edition, *Volume 2*, *Visual Cognition.*  MIT Press. Chapter 4, pp. 121-165.

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

Biederman, I., & Bar, M.  (2000).  Response to Hayward & Tarr (2000).  Vision research in press.

Biederman, I., & Bar, M.  (1999).  One-shot viewpoint invariance in matching novel objects. *Vision Research*, 39, 2885-2899.

Biederman, I., & Bar, M.  (1998).  Same-different matching of depth-rotated objects. *Investigative Ophthalmology & Visual Science*, 39, 1113.

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

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

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for 3D viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*. 19, 1162-1182.

Biederman, I., & Kalocsai, P.  (1997).  Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society London: Biological Sciences*, 352, 1203-1219.

Biederman, I., & Subramaniam, S.  (1997).  Predicting the shape similarity of objects without distinguishing viewpoint invariant properties (VIPs) or parts. *Investigative Ophthalmology & Visual Science*, 38, 998.

Biederman, I., Subramaniam, S., Bar, M., Kalocsai, P., and Fiser, J.  (1999).  Subordinate-level object classification reexamined.  *Psychological Research*, in press.

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.

Cooper, E. E., Subramaniam, S., & Biederman, I.  (1995).  Recognizing objects with an irregular part. *Investigative Ophthalmology & Visual Science*, 36, 473.

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

Edelman, S., & Bülthoff, H. H.  (1992).  Orientation dependence in the recognition of familiar and novel views of 3D objects.  *Vision Research*, 32, 2385-4000.

Esteky, H., and Tanaka, K.  (1998).  Effects of changes in aspect ratio of stimulus shape on responses of cells in the monkey inferotemporal cortex.  *Society for Neuroscience Abstr*acts, **24**, 899.

Hayward, W. G., & Tarr, M. J.  (2000). Differing views on views:  Comment on Biederman and Bar 1999.  *Vision Research*, in press.

Hayward, W. G., & Tarr, M. J.  (1997).  Testing conditions for viewpoint invariance in object recognition,  J*ournal of Experimental Psychology:  Human Perception and Performance,* 23, 1511-1521.

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

Hummel, J. E.,  & Stankiewicz, B. J. (1996a).  An architecture for rapid, hierarchical structural description.  In T. Inui and J. McClelland (Eds.). *Attention and Performance XVI: Information Integration in Perception and Communication* (pp. 93-121).  Cambridge, MA: MIT Press.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neuroscience*, *71*(3), 856-867.

Lades, M., Vortbrüggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R. P. and Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers, 42,* 300-311.

Logothetis, N. K.  (1999).  Paper presented at a workshop on Visual Object Recognition by Humans and Machines, Bad Homburg, Germany, May.

Logothetis, N. K., and Pauls, J., Bülthoff, H. H., & Poggio, T. (1994).  View-dependent object recognition by monkeys. *Current Biology*, 4, 401-414.

Mel, B. W.  (1997).  SEEMORE: Combining color, shape, and texture histogramming in a neurally-inspired approach to visual object recognition.  *Neural Computation*, 9, 777-804.

Nowak, L. G., & Bullier, J.  (1997).  The timing of information transfer in the visual system.  In *Extrastriate cortex,*  Cerebral Cortex vol.  12, Kaas J., Rockland K., & Peters A. (Eds.).

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

Riesenhuber, M., & Poggio, T.  (1999).  Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019-1025.

Shepard, R. N., and Cermak, G. W. (1973).  Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology*, 4, 351-377.

Tanaka, K.  (1993).  Neuronal mechanisms of object recognition.  *Science*, **262**, 685-688.

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

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

Tarr, M. J., Bülthoff, H. H., Zabinski, M., & Blanz, V.  (1997).  To what extent do unique parts influence recognition across viewpoint? *Psychological Science,* 8, 282-289.

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I.  (1998).  Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, 1, 275-277.

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

Ullman, S.  (1996). *High-Level Vision*.  Cambridge, MA:  MIT.

Vogels, R., Biederman, I., Bar, M, & Lorincz, A. (In press) Sensitivity of macaque temporal neurons to differences in view-invariant vs. metric properties of depth-rotated objects. *Journal of Cognitive Neuroscience.*

Wiscott, L., Fellous, J.-M., Krüger, N., & von der Malsburg, C.  (1997).  Face recognition by elastic bunch graph matching.  IEEE PAMI 19, 775-779.

Zerroug, M., & Nevatia, R.  (1996).  Volumetric descriptions from a single intensity image. *International Journal of Computer Vision*, 20, 11-42.