PERGAMON

Vision Research 39 (1999) 2885–2899

Vision Research

# One-shot viewpoint invariance in matching novel objects

Irving Biederman [a,*], Moshe Bar [b]

[a] *Department of Psychology, Neuroscience Program, Hedco Neuroscience Building, MC 2520, University of Southern California, Los Angeles, CA 90089-2520, USA*

[b] *Massachusetts General Hospital, NMR Center, Bldg 149, Rm 2301, 13th St. Charlestown, MA 02129, USA*

Received 12 January 1998; received in revised form 16 October 1998

**Abstract**

Humans often evidence little difficulty at recognizing objects from arbitrary orientations in depth. According to one class of theories, this competence is based on generalization from templates specified by metric properties (MPs), that were learned for the various orientations. An alternative class of theories assumes that non-accidental properties (NAPs) might be exploited so that even novel objects can be recognized under depth rotation. After scaling MP and NAP differences so that they were equally detectable when the objects were at the same orientation in depth, the present investigation assessed the effects of rotation on same-different judgments for matching novel objects. Judgments of a sequential pair of images of novel objects, when rendered from different viewpoints, revealed relatively low costs when the objects differed in a NAP of a single part, i.e. a geon. However, rotation dramatically reduced the detectability of MP differences to a level well below that expected by chance. NAPs offer a striking advantage over MPs for object classification and are therefore more likely to play a central role in the representation of objects. 

*Keywords:* Shape recognition; Shape representation; Object perception; Geons; Nonaccidental properties; Viewpoint invariance

## 1. Introduction

How are humans able to recognize an object when it is seen at a novel orientation in depth? Recent theorizing has coalesced around two general theoretical positions. According to view-based template theories, an object is represented as a set of 2D templates, one for each pose (e.g. Rock, & DiVita, 1987; Poggio & Edelman, 1990; Logothetis, Pauls, Bülthoff & Poggio, 1994; Tarr & Bülthoff, 1995). Matching with such templates is dependent on the metric properties (MPs) of the templates, e.g. length or aspect ratio, angles of intersections, and degrees of curvature. Such accounts hold that slight rotations in depth can be compensated by direct generalization from the templates, but greater orientation disparities incur a cost as a deliberative and slow operation, such as mental rotation, must be engaged to achieve recognition (e.g. Ullman, 1996; Tarr, Bülthoff, Zabinski & Volker, 1997). The primary observations motivating these theories derive from experiments in which highly similar, novel objects, such as a set of bent paper clips, show: (a) an increase in reaction times (RTs) and error rates in recognizing the objects when they are viewed at a different orientation in depth from that previously experienced; and (b) a reduction in rotation costs as new orientations in depth are learned (e.g. Tarr & Bülthoff, 1995).

However, several experiments have demonstrated that without prior familiarization little or no rotation costs can be manifested (Biederman & Gerhardstein, 1993; Logothetis et al., 1994). The critical feature of these demonstrations showing immediate viewpoint-invariance is the availability of *non-accidental properties* (NAPs) that differ from one object from another. These demonstrations reveal that only a single distinguishing NAP difference of a single part is sufficient to reduce or eliminate the rotation costs. The set of objects in such experiments could be regarded as highly similar subordinate level exemplars of a single basic-level class (Biederman, Subramaniam, Bar, Kalocsai & Fiser, 1999). NAPs are properties of objects that are relatively unaffected by rotation in depth, such as whether a given

* Corresponding author. Tel.: +1-213-7406094; fax: +1-213-7405687.

*E-mail address:* bieder@usc.edu (I. Biederman)



PII: S0042-6989(98)00309-5

contour is straight or curved, whether pairs of such contours are parallel or not, or whether the type of vertex formed by the co-termination of contours is an L, Y, or an arrow (Lowe, 1984)[1]. NAPs can be distinguished from MPs, which are affected by rotation in depth, such as an object's (or region's) aspect ratio or degree of curvature. Biederman and his associates (Biederman, 1987; Hummel & Biederman, 1992; Biederman & Gerhardstein, 1993) have posited a central role for NAPs in distinguishing among different kinds of simple parts, termed *geons*. According to this theory, an object is represented as a geon structural description, specifying the geons and the explicit relations between them (Biederman, 1987; Hummel & Biederman, 1992). As long as the same distinctive geons can be perceived at an equivalent level of discriminability, little or no cost in recognition across two different views should be apparent.

NAPs are not made explicit by metric template theories, although some view-based[2] theorists admit the use of NAPs with a known and restricted set of objects (e.g. Tarr & Bülthoff, 1995; Tarr et al., 1997). In general, however, view-based template theorists assign no special status to NAPs (as compared with MPs), in some cases specifically disavowing them (e.g. Edelman, 1995). The central issue under examination here is one of representation: Do NAPs enjoy a special status, allowing robust generalization to a new orientation in depth with a novel object, or are NAPs just like other possible stimulus measures, viz. MPs, that can be used to distinguish two shapes? Surprisingly, this critical issue has never been directly assessed with some scaling of the salience of the two types of image properties.

There is great interest in how objects can be recognized from a new orientation in depth because unlike 2D changes in an image's position, size, or orientation in the plane, which can readily be compensated by standard algorithms, rotation in depth alters the two-dimensional spatial components so standard techniques lose their capacity to distinguish one object from another at novel depth orientations. Yet people seem to be able to recognize familiar objects at new orientations in depth with little or no cost in recognition speed and accuracy (Biederman & Gerhardstein, 1993). Is this capacity based on prior experience with the particular objects or does it reflect a general capacity to exploit viewpoint-invariant properties, even for novel objects?

Biederman and Gerhardstein (1993) had demonstrated that the addition of a distinctive geon to a set of ten novel objects resembling bent paper clips (transforming them into charm bracelets), which would otherwise have been extremely difficult to recognize at unstudied orientations, resulted in nearly flat rotation functions ($\approx 5000°/s$) of reaction time against angle of rotation[3]. In their study, subjects first viewed an image of a charm bracelet at a particular pose, and then viewed images of 20 objects, either the target at rotations in depth of 0–60° or another charm bracelet. Tarr and Bülthoff (1995) protested this experiment, arguing that with a known target and a restricted set of alternatives, subjects could search for a distinctive NAP from the prespecified target. That human observers would exploit NAPs was exactly Biederman and Gerhardstein's interpretation, although they found no need to restrict their use to cases in which subjects had only a restricted set of possibilities. Rather, Biederman and Gerhardstein (1993, 1995) argued that descriptions composed of distinctive NAPs were the spontaneous and preferred representation at all levels of visual object classification.

The purpose of the present investigation into sequential object matching was two-fold. The first was a determination of whether depth invariance could be revealed spontaneously in a first-time encounter with an unfamiliar object under conditions in which subjects could not predict if, where, or how the second stimulus (S2) might differ from the first (S1). The second was a comparison of NAPs and MPs with respect to their utility for allowing recognition under depth rotation. The advantage of a matching task was that we could use novel stimuli without a familiarization stage. Moreover, we could manipulate precisely the information that the subjects could use to make their responses.

Two experiments are reported. In both experiments subjects judged whether a pair of stimuli were the same or different. When the stimuli were different they differed only in a single MP or NAP for a single part (or a single relation between the parts). This stimulus variation might correspond to highly similar members of the same basic-level class. In Experiment I (pure blocks), the pair of objects in each trial within a block were either always at the same orientation or always rotated in depth relative to each other. In Experiment II (mixed block), subjects attempted to detect the identical differ-

[1] Jacobs (1989) has argued that there are an infinite number of NAPs, once we consider the 2D projection of four or more non-coplanar features (points, lines, or curves). Recently, he has proposed that these NAPs that are used by human observers are restricted to those that can be defined with only a minimal amount of image information, three or fewer features (Jacobs, 1997). The present characterization of NAPs loses none of its generality if it is restricted to those of minimal complexity.

[2] The term view-based is an unfortunate terminology as all theories of object recognition, save those that assume ESP, must use the information in the image.

[3] The effects of rotation angle will be expressed in terms of degrees per second (°/s) rather than the more traditional seconds per degree to avoid the very small values that would result with the latter measure. Small effects of rotation angle are expressed as high values of degrees per second.

ences as in Experiment I, except that unrotated and rotated trials varied in random appearing fashion within a block.

## 2. Experiment I (pure blocks)

Subjects judged whether a pair of sequentially presented images of novel objects at an average depth rotation difference of 57°, were physically identical or not. There were several features of the design and procedure that we note.

1. *One-Shot*. To reduce the likelihood of employment of special purpose strategies that might suffice for a known set of stimuli, subjects were not given any training with the objects nor were they provided with any instructions as to the nature of the stimulus differences. Such strategies were claimed by Tarr and Bülthoff (1995) to be the basis of why Biederman and Gerhardstein's (1993) addition of a distinguishing geon to a charm bracelet would result in immediate viewpoint invariance. This aspect of the design also captures the ecological validity of the first-shot nature of the encounters that we have with many objects in our environment, such as a new model of a chair, where we have no prior instructional set as to what to look for.
2. *Equal detectability of NAP differences and MP differences from the same viewpoint*. When the objects differed (which they did on half the trials in most conditions) they could differ in a NAP or in an MP. A feature of the design was the calibration of the magnitude of NAP and MP differences so that they were equally detectable at the same unrotated orientation for a set of objects by human subjects. The effect of depth rotation could thus be studied unconfounded with the saliency of the shape differences when the objects were at the same orientation. Determining whether such saliency actually predicts rotation costs was also a goal of this effort.
3. *S1-S2 Position Shift*. The second object was always presented in a different position from that occupied by the first object, even when they were identical. The reason for this is that in a matching task, systems other than that required for object recognition can signal 'different' for any display change (Egeth, 1966). For example, Nowak and Bullier (1997) reported that IT cells have a transient response, originating in the magnocellular system, to any stimulus change. On a 'same' trial, with the second stimulus (S2) also at the same orientation as the first (S1), such a transient will be absent and the subject can respond 'same' immediately. However, a rotation of the object ('same' trial) or a change in its identity ('different' trial) will produce a transient as some previously occupied regions of the visual field will now be empty, and some previously empty regions will now contain some of the object's surfaces. By facilitating performance at 0° the transient (or its absence) provides an artifact that can produce apparent rotation costs. Biederman and Bar (1998) reported that shifting the stimuli even on same, 0° trials, greatly reduced the apparent rotation costs by increasing reaction times (RTs) and error rates for unrotated relative to rotated objects[4].
4. *Sufficiently long exposure durations*. In pilot testing it became clear that exposure durations of 200 ms for S1 and 100 ms for S2, that were sufficient for Biederman and Gerhardstein's (1993) line drawings, were insufficient for clear perception of rendered stimuli. We employed durations that were sufficiently long (400 ms for S1 and 300 ms for S2) to allow identification of the object's shape. The consequence of using too brief exposure durations in some prior experiments on the recognition of depth rotated objects is considered in Section 4.
5. *Pure and mixed designs*. In pilot testing it also was apparent that a mixed design in which rotated and unrotated trials would occur in random appearing fashion was more difficult than when all the trials in a block were unrotated or all rotated. To evaluate the effects of this variation, the investigation included both pure (Experiment I) and mixed (Experiment II) blocks of variation. This manipulation allowed us to assess the degree to which the calibrated equivalence between MPs and NAPs established under pure block conditions would also hold under conditions in which the subject could not predict the orientation of S2.

### 2.1. Method

#### 2.1.1. Subjects

A total of 40 subjects, age 18–41 years, participated either for payment or credit in Psychology courses at the University of Southern California. All had normal or corrected-to-normal vision. None were aware of the purpose of the experiment. The experiment had four groups, each with ten subjects.

[4] Because of the intervening mask, the transient in the present case would have to be one which was a function of the difference between S2 and an actively maintained representation of S1. Active maintenance would be necessary to avoid the disruption of the mask. Indeed, this is the subjective impression of what one is doing when performing this task. The transient may play a general role in the performance of same-different visual tasks, perhaps accounting for the paradoxical fast-same effect. By shifting the positions of the stimuli even on same trials, this effect is reduced in that transients would presumably always be signaling a difference. Consequently, the contribution of the translationally-invariant ventral cortical visual system, presumed to mediate object recognition, could be more accurately assessed. The possible impact of this transient on rotation costs in previous reports will be considered in Section 4.

### 2.1.2. Stimuli

A set of 12 rendered, simple (two-part), three-dimensional novel objects, each depicted at two orientations in depth, comprised the original set of objects. The images subtended an average visual angle of 8.2°. The objects were created on a Silicon Graphics Indigo2 work station, using the Inventor and Showcase Toolkits. A single, frontal light source was used. The images were displayed on a Macintosh 16'' color display, with a resolution of 832 × 624 pixels, and a refresh rate of 75 Hz. The stimulus presentation and response recording was controlled by a Macintosh Quadra 950.

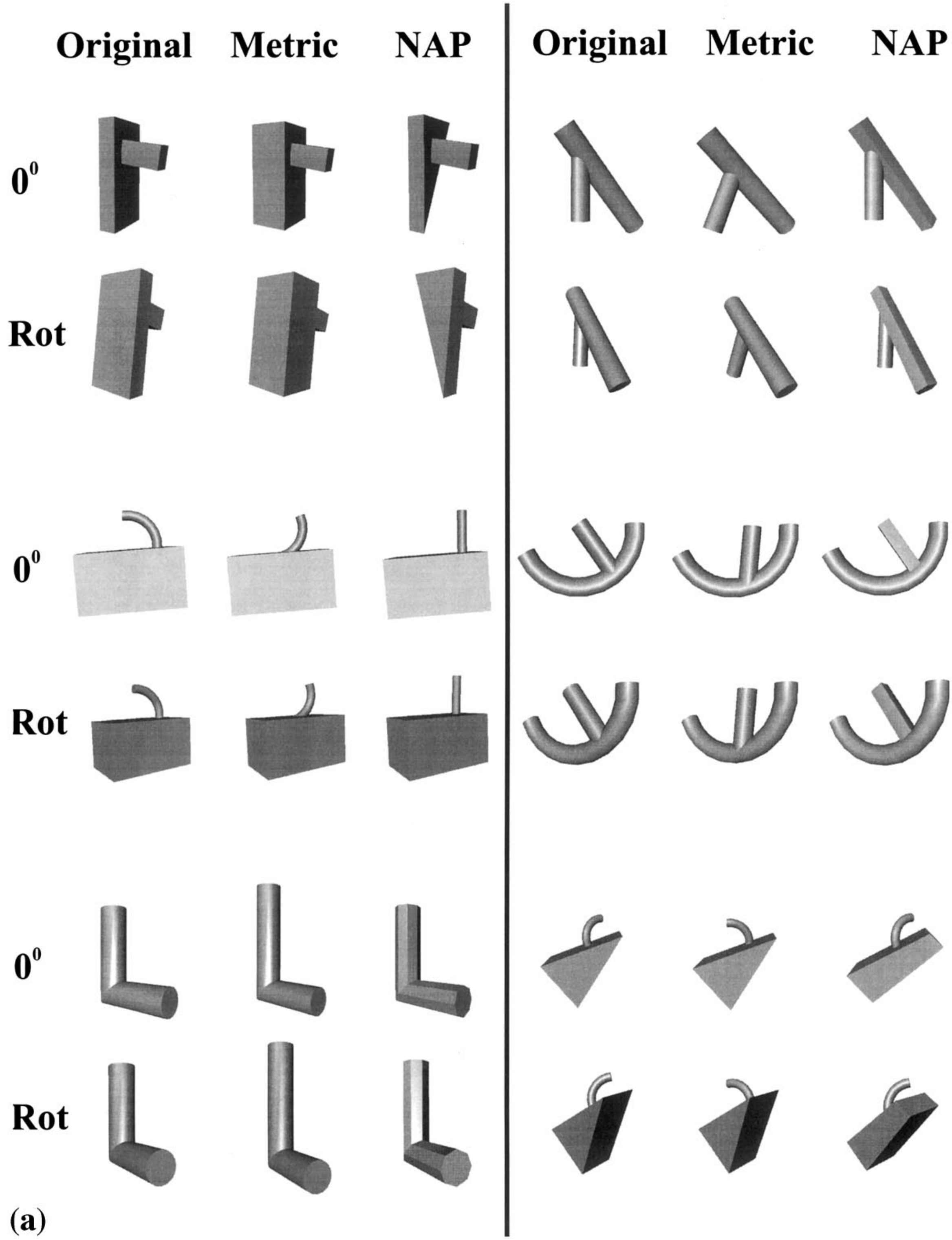


Fig. 1(a). *Caption opposite*.

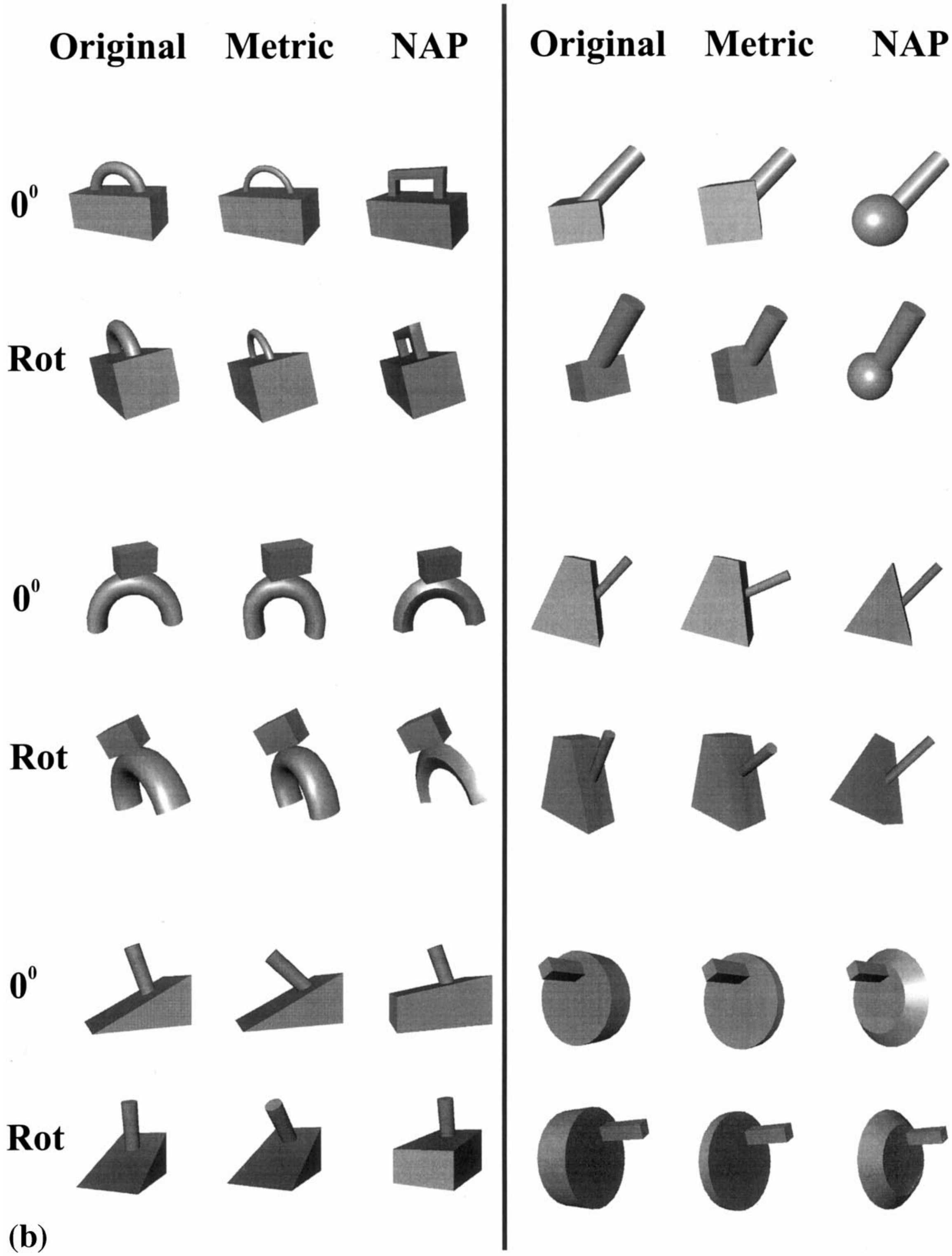


Fig. 1(b).

Fig. 1. The set of 12 three-dimensional novel objects. In the first column for each object is the original object, shown at the 0° (upper row) and rotated (mean = 57°, range 20–120°) orientations (lower row). The differences in surface lightness between the two orientations is a consequence of a single light source used in the rendering (which provided a potential cue as to the degree of rotation). On unrotated trials, S1 and S2 would be from the same row (and on same trials be the same object). On rotation trials, S1 would be from one row and S2 from the other. For different trials, one of the images would be an original from one row and the other image would be an MP or NAP from the other. The MP and NAP images were selected from a larger set of such changes to be, on average, equally detectable (from the original) at the 0° orientation. (a) The six objects whose MP and NAP differences from the original were equally detectable at both 0 and 57° (average) orientations. (b) The six objects whose metric and NAP differences from the original were equally detectable at only the 0° orientation. For these stimuli, the NAP differences were easier to detect, on average, than the metric differences.

One of the orientations for each object was arbitrarily designated as 0°. A single part of each Original object was modified such that it differed in either a MP, e.g. changing the aspect ratio of a part or varying its angle of intersection with another part, or a NAP of a part, e.g. changing a part's axis from curved to straight. These were produced subjectively to be of approximately equal salience. Three or four MP and NAP changes were made for each of the 12 original objects producing a total of 108 images. The changes that were made were often quite subtle in that the MP differences were limited to those that would not change a qualitative relation between the two parts, e.g. change their relative size, and the VIP difference had to be selected to be approximately equivalent in salience to the metric difference.

From the performance of the unrotated 0° group, a set of MP and NAP differences that were, on average, equally detectable yielded a set of 36 objects (an original, an MP, and a NAP for each of the 12 original objects). These are shown in the upper row for each of the original objects in Fig. 1(a, b). For eight of the 12 original stimuli, the MP and NAP changes were made to the same part. Each of these objects was rotated around the vertical an average of 57° (range 20–120°). These are shown in the second row for each of the objects in Fig. 1(a, b). The angles were selected to be the rotations that would still allow the same parts to be identifiable in the two orientations, although some near accidents were produced and the different orientations often differed greatly in their reflectance characteristics.

### *2.1.3. Design*

Four groups of subjects were assigned to three conditions as follows:

1. *Unrotated 0° condition*. Ten subjects performed 143 trials in this condition in which S1 and S2 were both depicted at the same 0° unrotated orientation. On 57 of the trials (approximately 40%) the stimuli were identical. On 86 (60%) of the trials, the stimuli differed in either a NAP or an MP (with equal probability). For all three conditions, the trial order was reversed for every other subject so that the mean serial position of every trial type, defined by a particular S1… > S2 sequence, was equal across subjects. One of the groups of ten subjects who performed the condition rotation A (described below), also performed this task. This group thus provided an opportunity for replication of the original equivalence selection as well as a within-subjects comparison of performance with rotated and unrotated objects.
2. *Unrotated 57° condition*. Subjects judged whether a sequential pair of objects was the same or different when both objects were shown at the same orientation, which was rotated an average of 57° from the 0° orientation. Each subject performed 192 trials, 96 same and 96 where there was an MP or NAP difference (equally likely to occur). The number of MP and NAP different trials was equally distributed over the experiment. Within the first 96 trials, subjects viewed particular MP and NAP different trials for a given object type twice, once when the original object came first and once when the changed object came first. This allowed an analysis of the first time a subject had a particular trial type for a given object. The same trials were repeated in the second half of the subject's session but in different orders. This condition allowed a subset of six of the 12 stimuli to be selected for which the MP and NAP differences were equally detectable both at the unrotated 57° orientation, as well as at the unrotated 0° orientation.
3. *Rotation condition*. Subjects judged whether a pair of objects, which were always depicted at different orientations, were the same or different. All the stimuli shown in Fig. 1 were used here, which included the subset of six MP and NAP differences that were equally detectable at both the 0 and 57° orientations (and the six which were only equivalent at 0°). Each subject performed 192 trials in this condition. One of the two images was always an original, the other could be either an original, to be responded with the 'same' response, or an MP or NAP change, to be responded with the 'different' response. Half the trials were 'same' and half 'different'. The 96 different trials were equally divided into 48 MP and 48 NAP differences. For each forward order of stimuli, the reversed order was also included for a different subject. In addition, the trial order was reversed for every other subject so that the mean serial position of every trial type, defined by a particular S1… > S2 sequence was equal across subjects. Two groups of ten subjects each performed this task. In the rotation A group, ten subjects performed this task as well as the 143 trials of the 0–0° unrotated condition. Five performed the rotation condition first and five the unrotated condition first. another group of ten subjects, rotation B, performed only this task, providing a replication of the rotation condition.

### *2.1.4. Procedure*

The sequence of events on a trial is illustrated in Fig. 2. Following a press of the mouse button, a fixation dot appeared for 500 ms, followed by a 400 ms presentation of the object, which was then immediately followed by a mask consisting of a combination of different gray-level objects presented for 500 ms. A second object image was then presented for 300 ms, followed by a second 500 ms mask. The second stimulus was translated randomly to one of the remaining eight possible

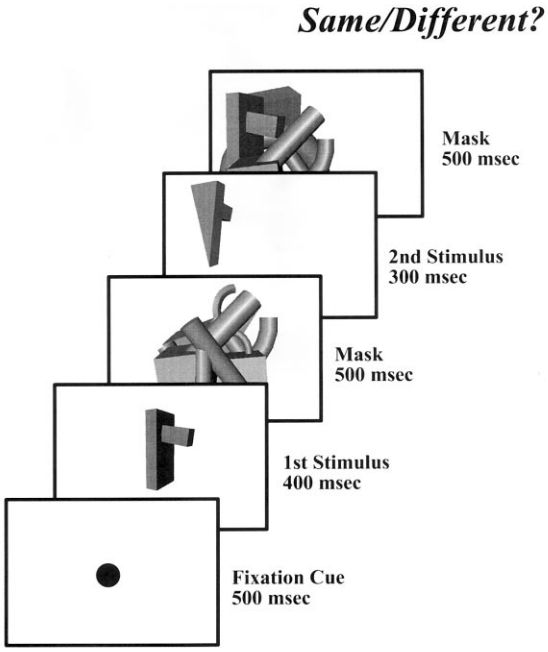


Fig. 2. Sequence of events on an experimental trial. An illustration of a NAP different, rotation trial. Note the shift of the second object to the upper left relative to the position of the first object.

positions on the screen, specified by a 3 × 3 matrix with adjacent horizontal or vertical centers separated by 6.8°. Thus, the second image could be above or below, and/or to the right or to the left of the first image which was always centered. To reduce habituation to the masking (Intraub, 1984), four different masks, randomly selected, were used.

Subjects were instructed to ignore the intervening mask, and when the second image appeared, to press as quickly as possible a microswitch labeled 'same' if the object depicted in the second image was identical to the first, and a microswitch labeled 'different' if they were images of different objects (differing in a MP or a NAP). Subjects could not anticipate whether there was going to be a stimulus change or, if there was a change, whether it would be of an MP or a NAP, or what particular part would undergo the change. No feedback was provided during the experiment.

## 2.2. Results

One image with an MP change and one with a NAP change were chosen for each original object in the unrotated 0° group so that their mean RTs and error rates for detecting the changes were equivalent. These images are shown in the upper row for each object in Fig. 1. The mean RTs and error rates for detecting the MP and NAP changes were 774 ms and an error rate of 20.0% (the 0° data are shown in Fig. 4). This equivalence was not a consequence of a few outliers; there was substantial overlap in the distribution of MP and NAP RTs and error rates. The relatively high error rates in this unrotated condition was likely a consequence of the subtle differences to be detected and, perhaps, some difficulty in suppressing a different response to shifted stimuli or in recovering from such suppression when the stimuli really were different. From these 36 images used in the Nonrotation 57° group, a subset of 18 images (six original stimuli and the corresponding MP and NAP variations) were selected in which the MP and NAP differences were equally detectable also from the 57°orientation. These are the six objects shown in Fig. 1(a). This additional selection was done because the full set of 12 MP and NAP differences were not equally detectable at the 57° average orientation: The NAPs for the full set were considerably more detectable than the MPs at that orientation.

Fig. 3 shows the results for the rotation A group (where the subjects had both the unrotated 0° condition as well as the rotation conditions in separate blocks) for

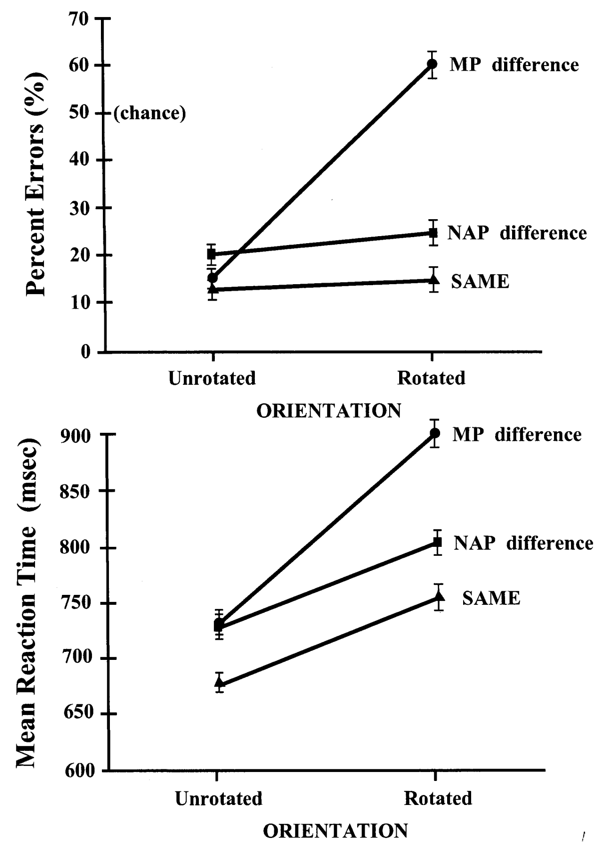


Fig. 3. Experiment 1, within subjects, pure blocks. Mean error rates (upper panel) and mean correct RTs (lower panel) for ten subjects (rotation A group) as a function of differences in orientation and the type of change (MP difference, NAP difference, or same). Data are for the six objects whose MP and NAP differences were equally detectable at both 0 and 57° on their first occurrence of a given S1-S2 sequence. Subjects in this group performed both the 0° unrotated and rotation conditions, and the data points in unrotated and rotated are therefore from the same subjects. RTs greater than 2000 ms were counted as errors. Error bars are the S.E. after removal of between-subjects variance.

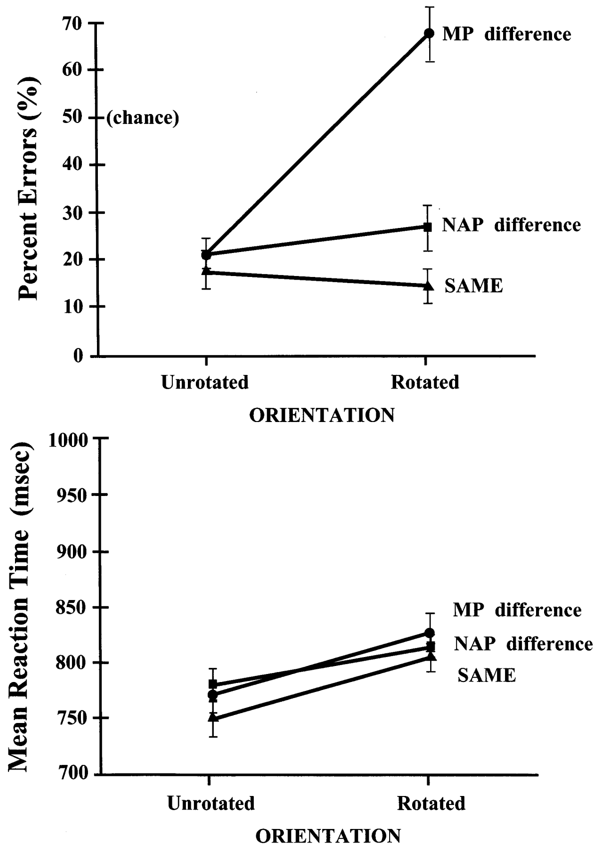


Fig. 4. Experiment 1, between subjects, pure blocks. Mean error rates (upper panel) and mean correct RTs (lower panel) as a function of differences in orientation and the type of change (MP difference, NAP difference, or same) for three groups of ten subjects each. Data are for the six objects whose MP and NAP differences were equally detectable at both 0° and 57°. One group of subjects performed only the 0° unrotated condition, one group only the 57° unrotated condition, and one the rotated B condition. RTs greater than 2000 ms were counted as errors. The error bars are the between-subjects S.E.

the six objects whose MP and NAP differences were equally detectable at both orientations. The data are for the first time a given stimulus sequence was experienced. Because all the unrotated sequences were only presented once, general practice effects would favor the unrotated over the rotated condition. The mean serial position for the first time a stimulus sequence was shown (ignoring ordering of S1 and S2) was 71 for the unrotated condition and 48 for the rotated condition. There was only a slight effect of rotation in depth on the detection of NAP differences: an increase of 3.3% in error rates and 75 ms in mean correct RTs. In contrast, rotation increased error rates by 46.2% for the MP group, to a level (59.5%) that was below chance. With such high error rates, the interpretations of RTs would be meaningless but we note that there was also an increase of 172 ms in RTs when matching MP differing stimuli at different orientations, precluding a speed-for-accuracy tradeoff as an account of the higher error rates for detecting differences in MPs compared to detecting differences in NAPs.

The pattern of results for all the trials in the Rotation condition, shown in Appendix A, was virtually identical to that shown in Fig. 3. (The unrotated trials are the same values because those sequences were only experienced once.) The detection of rotated MP differences did not benefit from the prior trials. Error rates increased by 5.9% compared to the value on the first trial whereas there were slight reductions in the error rates for detecting NAP differences and same responses. The costs of rotation on RTs declined in all conditions but were still markedly higher in the MP condition (153 ms increase) compared to the NAP condition (36 ms increase). The data for the full set of 12 objects (calibrated only at 0°) was also highly similar to that of the restricted set of six objects (calibrated at both 0 and 57°) except that there were no rotation costs on the RTs for detecting NAP differences (actually a slight decline of 15 ms) or judging that the objects were the same. These data are shown in Appendix B. In general there was little systematic change in performance over trial blocks

Fig. 4 shows the between-subjects results for the first trial that a given S1-S2 sequence was experienced for the six objects, calibrated at both 0 and 57°, for the unrotated 0°, unrotated 57°, and rotated B groups. The pattern is highly similar to the pattern found for the within-subjects comparisons shown in Fig. 3. The equivalence in performance in detecting MP and NAP differences for the unrotated groups gave way to markedly inferior MP detection performance for the rotated groups, providing the same qualitative picture as that shown in Fig. 3: only a small effect of rotation on detecting the NAP differences but massive costs for detection of MP differences.

The between-groups results for all the trials for the six objects is shown in Appendix C and the results for the full set of 12 object types is shown in Appendix D. Both tables reveal a highly similar pattern of results as those shown in Fig. 4 (and Fig. 3), with respect to the error data. Rotation produced almost no effect on error rates for detection of NAP differences but rendered performance below chance for MP detection. RTs for detection of NAPs in these conditions was only slightly affected by rotation but these costs were always much smaller than they were for detection of MPs.

Did the scaling at the unrotated value actually matter in producing the rotation costs? To address this question we correlated the average unrotated error rates with the error rates under rotation for each stimulus. Table 1 gives the correlations for error rates and RTs separately for the within- and between-groups comparisons, for the original ('Same' response), MP, and NAP ('different' responses) conditions and for both the restricted set of stimuli (calibrated at both orientations) and those calibrated only at 0° orientation. It is clear

Table 1
Pearson product moment correlations between error rates for same–different judgments for unrotated stimuli and rotated stimuli

| Condition (response) | Six objects (df = 4) | 12 objects (df = 10) |
|---|---|---|
| *Within* | | |
| Original (same) | 0.34 | 0.76** |
| MP (different) | 0.77[a] | 0.54[a] |
| NAP (different) | 0.48 | 0.46 |
| *Between* | | |
| Original (same) | −0.19 | 0.47 |
| MP (different) | 0.89* | 0.61* |
| NAP (different | 0.69 | 0.48 |

[a] Two-tailed $\alpha$ levels = 0.10.
* Two-tailed $\alpha$ levels = 0.05.
** Two-tailed $\alpha$ levels = 0.01.

that there was at least a moderate positive relation between the difficulty in judging a given object as being same or different when it was unrotated and when it was rotated. (Only one of the 12 correlations was negative, and that only slightly so.) For all four comparisons, the correlations were stronger for MPs than for NAPs. The tabled values undoubtedly underestimate the actual strength of the relations because of variability in the estimates of unrotated and rotated performance and the restrictions of range required to produce equal mean detectability of MPs and NAPs at unrotated orientations.

## 3. Experiment II (mixed blocks)

Experiment I was run with pure blocks in which the object shown in S1 and S2 was either always depicted at the same orientation or always depicted at a different orientation. The absence of orientation uncertainty in Experiment I was designed to provide less demanding conditions for detecting shape differences. Experiment II was designed to provide a closer approximation to many real world encounters with objects in which there is some uncertainty as to an object's orientation before it is experienced. Under these conditions, subjects could not reliably anticipate whether S2 would be rotated or unrotated. Would the equivalence of unrotated MP and NAP differences established in the pure blocks be obtained under these more varied conditions?

### 3.1. Method

A total of 40 subjects participated in this experiment. None had participated in Experiment I. The design was similar to those of the previous experiments, except that on each trial S1 and S2 could either be depicted from the same orientation or from different orientations. On half of the 192 trials, S1 and S2 were depicted at the same orientation; in the other half they were at different orientations. On half the trials for each rotation type, S1 and S2 were the same; on the other half they differed in an MP or a NAP (with equal frequency). Half the unrotated trials were shown from the 0° orientation; the other half from the 57° orientation. The distribution of trials was as follows: 48 were 0–0° (12 MP, 12 NAP, and 24 same), 48 were 57–57° (12 MP, 12 NAP, and 24 same), 48 were different trials under rotation (24 MP and 24 NAP, with each pair appearing twice by switching the orientations of S1 and S2), and 48 were same trials under rotation (24 originals, and 12 trials each of the same MP and NAP objects as S1-S2).

### 3.2. Results

Fig. 5 shows the data for the six stimuli that were calibrated from both starting orientations in Experiment I. RTs for all conditions were dramatically higher than they were for the corresponding pure block tasks

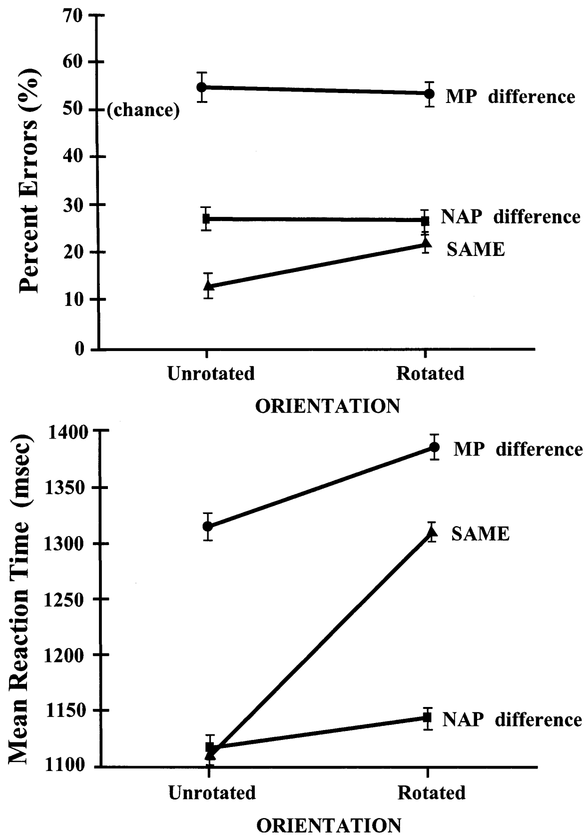


Fig. 5. Experiment 2, within subjects, mixed blocks. Mean error rates (upper panel) and mean correct RTs (lower panel) for 40 subjects as a function of differences in orientation and the type of change (MP difference, NAP difference, or same). Data are for the six objects whose MP and NAP differences were equally detectable at both 0 and 57°. Subjects in this group performed both the 0° unrotated, 57° unrotated, and rotation conditions, and the data points in unrotated and rotated are therefore from the same subjects. RTs greater than 2000 ms were counted as errors. Error bars are the S.E. after removal of between-subjects variance.

shown in Fig. 3. In spite of the increased difficulty of the mixed task, the detection of NAP differences remained largely unaffected by rotation in depth. Because the detection of unrotated MP differences was below chance, the lack of an effect of rotation is largely uninterpretable.

Not only was the mixed block task considerably more difficult than the pure block tasks of Experiment I, it was also the case that the equivalence scaling for detecting MP and NAP differences in unrotated objects, established in Experiment I, no longer held. Detection of differences in MPs were now far more difficult than detection of differences in NAPs for unrotated objects. Both the increased general difficulty in performing this task, as well as the loss of the equivalence scaling, is likely a manifestation of an increased cognitive (attentional?) load produced by the uncertainty in orientation of S2 following perception of S1. The data for the full set of 12 stimuli (Appendix E) are almost identical to the data shown in Fig. 5.

## 4. Discussion

The results provided strong support for an immediate and striking advantage for the employment of NAPs over MPs for stimuli differing in a single part. When the objects differed in a NAP of a single part only small effects of depth rotation on error rates or RTs were evident. In contrast, error rates for detecting differences in MPs of objects at different orientations in depth were dramatically higher than those for NAPs—to a level below that expected by chance, accompanied by a marked increase in RTs. These results provide a challenge to view-based, template theories that do not assign a privileged status to NAPs. The results also suggest that previous reports of high rotation costs (e.g. Poggio & Edelman, 1990; Logothetis et al., 1994; Tarr & Bülthoff, 1995) with novel objects such as bent paperclips were a consequence not so much of the unfamiliarity of the stimuli, but of the absence of NAPs that could be employed to distinguish the objects.

That the detection of MP differences manifested greater costs than the detection of NAP differences under the additional load of a mixed task in Experiment II suggests that the detection of MPs, relative to the detection of NAPs, is a more attention-demanding activity and may be less likely, therefore, to be spontaneously deployed in object recognition.

In the current experiments modest NAP changes were employed in that the relations between the parts remained intact, and the overall variation was calibrated to be equivalent to an MP change. These NAP differences more closely resemble subordinate-level distinctions among highly similar exemplars. Given that near viewpoint invariance was achieved with the modest NAP variations in the present investigation, we would expect it to be even more strongly evident when distinguishing NAP variations that approximated what is encountered in the real world when distinguishing among objects at a basic-level or more typical subordinate level distinctions, such as that between a round table and a square one. We would expect, for example, that there would be little or no rotation costs if subjects had to distinguish one original object from another original object in Fig. 1.

In noting that immediate viewpoint invariance is possible, we do not wish to convey the impression that there is no role for learning of different viewpoints. Our conclusions are limited to those views where the same geon structural descriptions (GSDs) can be activated. Obviously, from the view of a front of a house one would have to learn what the back looked like. To the extent that there are different NAPs in different views, new representations, i.e. new GSDs, would have to be developed for those views and invariance would not be expected nor is it found[5]. Although the initial activation of a representation of an object may be invariant to viewing variables, such as position, size, and pose, these viewpoint variables are remembered nonetheless (Biederman & Cooper, 1991a,b, 1992; Cooper, Biederman & Hummel, 1992). Thus, for example, although there is no effect on name priming of mirror reflection of an image of an object, subjects have good explicit memory as to the object's orientation. Biederman and Cooper (1992) speculated that the priming and, presumably, the same-different matching in the present experiment[6], was mediated by a representation that was invariant to position, orientation, and size, but explicit memory accessed a representation that bound these viewpoint variables to the specific shape of the object to produce an episode. Under difficult viewing or judgment conditions with a restricted set of stimuli, subjects could employ such episodic information to make their response. If the table is on the right and the chair is on the left, a very brief flash of an object on the right could be called a table.

[5] Biederman's (Biederman, 1987, 1988) 'componential recovery principle' held that the similarity between two views of an object (produced by rotation or occlusion or removal or addition of parts) is a positive function of the common geons in the two views and a negative function of the geons present in one view and not the other. Biederman and Gerhardstein (1993) also noted that a resolution function was required to determine the degree to which a geon's visibility might be affected by variations in viewing conditions. As discussed later, mirror reflection offers a strong test of this principle without the requirements to assess the perceptibility of the geons in that reflection changes the object's view-based shape but not its GSD.

[6] Stankiewicz et al. (1998) have argued that attention is required to create a view-invariant representation in matching. There can be little doubt that the subjects were attending to S1.

We consider now several aspects and possible objections to the present account.

'*The novel objects are not really novel in that the parts are familiar*.' This is true. It is also true of the narrow, elongated cylinders that comprise bent paperclips and other such stimuli in which immediate viewpoint invariance is not manifested. People have seen pipes, not to mention bent paperclips, joined end-to-end at various angles so the parts and relations comprising bent paperclip stimuli certainly have been encountered in our lives. Nonetheless, immediate viewpoint invariance is impossible with such stimuli (e.g. Tarr et al., 1997), but was readily obtained with our novel objects. The critical factor would appear to be not the familiarity of the stimuli or the stimulus parts, but the requirement that the information be (a) nonaccidental, and (b) distinctive. These are some of the requirements for viewpoint invariance argued by Biederman and Gerhardstein (1993).

'*View-based models can include the right features so the account presented here does not challenge such template theories*.' Undoubtedly, a template theory can be formulated that does make explicit NAPs. If such a theory also made explicit the part structure of the object it would start to resemble geon theory, with the sole distinction being that in a template representation the relations among the parts are implicit in a coordinate space whereas in geon theory they are expressed explicitly in the units of a structural description. The analyses of experimental results by template theorists, to date, has ignored NAPs (e.g. Poggio & Edelman, 1990) or attempted to disavow the pivotal role played by such information (e.g. Edelman, 1995; Tarr & Bülthoff, 1995; Haywood & Tarr, 1997; Tarr et al., 1997; Tarr, Williams, Hayward & Gauthier, 1998).

'*Same-different matching tasks may not generalize to object identification tasks*.' Matching tasks have typically employed repeated trials with the same small set of stimuli and, consequently, subjects can evolve strategies to restrict their processing only to those aspects of the stimuli that are relevant for that particular task. These strategies may not be operative when people are identifying objects. Our particular implementation of the matching task, however, had position and feature uncertainty for novel objects. Consequently, subjects could not know where to look or how the stimuli might differ, and they could not capitalize on the absence of a transient for same responses. Under such conditions, subjects would be induced to represent the complete first stimulus—they could not know which part, if any, would change nor how it might change—in a manner that would allow them to achieve rotational invariance, if possible. This is precisely the behavior that we were interested in assessing.

A task in which subjects are trained with arbitrary names for a particular pose of an object (Tarr et al., 1998, Experiment 3; Haywood & Tarr, 1997, Experiment 2), particularly if the distinguishing information is difficult to discriminate, is problematic insofar as pose is part of what is learned and, potentially, used. If one is trained to name a particular object that is, for example, facing left, the subject may well associate the name to that shape *and* its orientation, despite instructions to ignore pose. Obviously, we have learned to distinguish a *p* from a *q* and a *b* from a *d* using only pose information (which is, perhaps, why these letters offer maximal confusion to the novice reader). Basic-level object naming does not suffer from this ambiguity in the task demands.

'*Other experiments obtain rotation costs for stimuli differing in NAPs*.' A number of studies have documented rotation costs where the rotations occluded some geons and revealed others or produced accidental or near accidental views (e.g. Humphrey & Kahn, 1992; Srinivas, 1993). Michael Tarr and his associates (Haywood & Tarr, 1997; Tarr et al., 1997, 1998) have recently reported rotation costs when accidental views were, presumably, controlled. The magnitude of these costs were small, as they were in the Biederman and Gerhardstein (1993) and the present experiment, relative to the rotation costs incurred with stimuli that do not differ in GSDs. For example, Tarr et al. (1998) studied the recognition of rendered single geons adapted from Biederman and Gerhardstein's (1993) experiment 4 with line drawings. (One of Tarr et al.'s 1998, nine experiments did use line drawings.) If the slope of the plot of reaction time against rotation angle (from 0 to 90°) is expressed in degrees per second, then the rotation rates for these stimuli ranged from approximately 750°/s for a naming task to 3600°/s for a match-to-sample task with a go/no-go response. For some experiments, Biederman and Gerhardstein reported flat functions or orientation costs of only 5000°/s.

*Resolution artifacts*. Do these slopes, as shallow as they are, represent fundamental view-dependence as would be expected from, for example, mental rotation or the extrapolation or interpolation of templates, as Tarr and Bülthoff (1995) have argued, or do they represent variations in extracting GSDs at different rotation angles? Although the possibility of a template-like representation cannot be definitively ruled out to account for some of these rotation costs, there are a number of factors other than template mismatching that could have contributed to these costs. Despite the attempt at avoidance of accidental views, many of the views in Tarr et al. (1998) were, in fact, near accidents that required, for example, determination of whether a single small contour was straight or slightly curved to distinguish one object from another[7].

Before considering the details of how confounds of resolution and other factors might have artifactually produced rotation costs in these other studies, we note that there is independent evidence that resolution variations may be sufficient to produce the observed rotation costs. Curiously, most such experiments reporting significant slopes with (presumably) distinctive NAPs have studied relatively small rotation angles, up to 90° and, in some cases, only to about 30° (Haywood & Tarr, 1997). From a view-based perspective, a rotation of 180° or mirror reflection of a bilaterally symmetrical object would be expected to produce enormous rotation costs, relative to these slight rotation angles. The opposite, however, occurs. Mirror reflections incur *no* cost in priming in people (Biederman & Cooper, 1991a,b; Stankiewicz, Hummel & Cooper, 1998) and monkeys (Logothetis et al., 1994)[8]. Reflected images present an ideal case with which to evaluate geon structural descriptions insofar as GSDs do not distinguish such reflections and the problem of geon resolution is precisely controlled. The lack of any effect of mirror reflection is completely consistent with the Hummel and Biederman (1992) implementation of geon theory. Even rotations short of 180° show a benefit. Biederman and Gerhardstein (1993) (Experiment 1) found no effect of a 135° rotation on the priming of object naming.

The procedures of the present investigation were designed to minimize the artifacts that can lead to apparent rotation costs with stimuli that differ in NAPs. The essential point here is that rotation in depth tends to produce drastic changes in the 2D image that can differentially affect the perceptibility of the parts. Rendered images, as compared to line drawings, typically have lower contrast and illumination and shadow contours that can increase the difficulty of determining the orientation and depth discontinuities important for resolving the geons. Such resolution difficulties characterize the object images in the Tarr et al. (1997) and Haywood and Tarr (1997) experiments. As an object is rotated in depth, these effects can vary for different geons. (Resolution problems caused by rendering are apparent to anyone who has attempted to assemble equipment from similar parts when the parts are displayed as photos rather than depicted as line drawings.) Biederman and Bar (1998) showed that by increasing stimulus presentation durations the rotation costs for rendered stimuli (geon charm bracelets similar to that of Tarr et al. (1997)) were markedly reduced.

*Transient artifact*. In same-different matching tasks, transients are produced when rotated stimuli no longer occupy the same regions of the screen so the absence of a transient is a reliable cue that unrotated stimuli are the same. Shifting all stimuli reduces the impact of the transient, even when they are not rotated. The shift increases the difficulty of the 0°-rotation condition relative to the positive rotation conditions, thus producing lower rotation costs. Biederman and Bar argued that these transient shifts were the reason why, in the Haywood and Tarr (1997) and Tarr et al. (1997) studies, a rotation from 0° to a slight angle, say 30°, produced greater costs than rotations from greater angles, say from 60 to 90°. The opposite would be expected from the template extrapolation/mental rotation routines argued by Tarr (1995).

The present results are thus consistent with the results of Biederman and Gerhardstein (1993) with line drawn charm bracelets. The line drawings provide better contrast for perception of the orientation and depth discontinuities required for the perception of the geons. Only by degrading the distinctive NAPs, through rendering and using brief exposures, or allowing near accidents, or allowing artifactual increases on rotated stimuli, e.g. the transient artifact, are rotation costs apparent with such stimuli. Even under such conditions, the rotation costs are always modest relative to stimuli that cannot be distinguished by NAPs[9].

[7] Evidence to this point can be obtained from a comparison of performance of Tarr et al.'s (1998) experiments and that of Biederman and Gerhardstein's Experiment 4. Both used a go no-go, match-to-sample task with single geons. In the Biederman and Gerhardstein experiment, although most of the distractors never elicited a false alarm, some had false alarm rates of 60–100%. Their subjects were induced to respond quickly and evidenced almost no rotation costs, but a 15–20% false alarm rate. In contrast, Tarr et al.'s (1998) subjects responded far more slowly but with a false alarm rate of only 5% and a (modest) slope of 2250°/s. It is likely that subjects in the Tarr et al. experiments were taking the time to resolve the small differences in contour needed to reject a near distractor.

[8] If one assumes that the object is bilaterally symmetrical then an algorithm developed by Vetter and Poggio (1994) can match mirror reflected images without a costly normalization procedure. However the application of such an algorithm would produce no costs for rotation to *any* angles.

[9] Cohen and Kubovy (1993) argued for a criterion of 1000 °/s as the fastest rate at which 2 D mental rotation can be performed. Rates greater than this value presumably could not be a consequence of mental rotation. Tarr et al. (1997) have taken this figure as a criterion of mental rotation so that rates below 1000°/s are interpreted as possibly being matched through mental rotation. We do not accept this criterion. Calculated rotation rates differ greatly for different sets of stimuli (Hochberg & Gellman, 1977) so, if mental rotation does exist, it should be assessed separately for each set of stimuli. With objects resembling bent paper clips initial error rates are so high (often below chance) that it is virtually impossible to determine a latency measure but the value would surely be well under 1000°/s. With a distinctive geon, Tarr et al. (1997) did obtain a rate under 1000°/s but the various problems noted earlier with this study (viz. potential poor resolution of the geons, the transient artifact, short rotation intervals) raise questions about whether it was mental rotation that was basis of the rotation costs. Incidentally, the critical attribute in Hochberg and Gellman's study serving to produce low rotation costs was the presence of a salient landmark of the kind that would be readily expressed by GSDs.

### 4.1. Do distinctive GSDs offer a benefit in reducing rotation costs?

There would seem to be no question that when a set of objects lack distinctive GSDs, the ability to recognize them from arbitrary viewpoints would be much worse than when distinctive GSDs are present. Thus Rock and DiVita's (1987) subjects were at near chance levels in recognizing which of two smooth complex novel wire objects they had seen previously. Anyone who has tried to recognize a rotated bent paper clip from among other bent paper clip distractors, of the kind studied by Edelman and Bülthoff (1992), quickly realizes the extraordinary difficulty in performing such a task compared to charm bracelets.

### 4.2. Geon structural descriptions as an account of object recognition

Are GSDs an appropriate representation to characterize this advantage of stimuli that differ in geons and relations compared to those that do not so differ? Tarr et al. (1997) performed a same–different matching task with rendered versions of the Biederman and Gerhardstein's (1993) charm bracelets and a comparable set of paper clips. As would be expected from the previous discussion, the charm bracelets were far easier to recognize under rotation: At 90° the d′ for the charm bracelets with a single distinguishing geon was approximately 3.0; for the paper clips it was 0.5. Tarr et al. (1997) also included charm bracelets with three or five different geons, sampled from a set of ten geons. The additional geons reduced performance so that with five different geons the d′ at 90° was 2.0 (still markedly greater than the d′ for the paper clips). Tarr et al. interpreted this last result as evidence against GSDs, insofar as the additional geons did not facilitate performance. But this interpretation is mistaken. Hummel and Biederman (1992) had argued that their network would not be able to distinguish a linear array of three geons, much less five. Part of the reason is that with single place predicate relations of the kind assumed by Hummel and Biederman (e.g. a cylinder side-of another geon), the inner orders of geons are not distinguished. More generally, with the identical set of side-of relations for all geons in all stimuli and ten five-geon subsets of the same ten geons, the vectors describing the different objects would be highly similar, thus reducing their discriminability. The consequence of this is that one would have to employ complex rules to distinguish the stimuli, e.g. if the middle geon is a cylinder and one of the end geons is a wedge then if the geon on the other side of the cylinder is a brick it is object A but if that geon is a cone it is object B. Biederman and Gerhardstein (1993) explicitly argued that stimuli from such sets are not distinguishable by GSDs. In contrast, the NAP differences in the present experiment are well expressed by GSDs.

When investigating same-different matching of a set of bent paper clips (of the kind studied by Tarr et al., 1997), Biederman and Bar (1998) discovered striking differences in the miss and false alarm rates at the identical rotation angles for individual pairs of images. Put simply, if the images projected by the first and second stimuli differed in a qualitative feature, e.g. an arrow vertex for one and a near linear array for the other, then the subject tended to respond different, producing high miss rates ($>50\%$) when the objects were the same and relatively low false alarm rates when the objects were actually different (27%). When the images did not differ in a qualitative feature, then the subjects tended to respond same, producing high false alarm rates (as high as 88%) when the stimuli were different and low miss rates when they were the same (as low as 4.6%). GSDs would appear to be an apt representation not only for charm bracelets and the objects used in the present investigation, but for these paper clip stimuli as well, a point argued by Biederman and Gerhardstein (1993, 1995). Confirmation of this conclusion can be found in the observations of view-based proponents (in the restrictive sense of view-based) themselves. In training monkeys to respond to a particular object at varied orientations, Logothetis et al. (1994) noted:

> …when the wire-like objects had prominent characteristics, such as one or more sharp angles or a closure, the monkeys were able to perform in a view-invariant fashion, despite the distinct differences between the two-dimensional patterns formed by different views… the animals easily learned to generalize recognition to all novel news of basic objects [such as a teapot or spaceship]… the objects were considered basic because of their largely different shape from the distractors... The monkeys had never seen these objects before… So, their remarkable performance may be the result of quickly learning… some characteristic features of the objects, for instance, the lid's knob or the handle of the teapot, or some relationship between such features and a simple geometrical shape, endowed with an axis of symmetry (p. 411).

## 5. Conclusions

The main conclusions of this investigation are that: (a) NAP differences are dramatically more detectable than differences in MPs in the matching of depth-rotated objects, even when the two types of differences were made to be equally detectable at 0° orientation disparity; and (b) When novel objects differ in a NAP, immediate viewpoint invariance over rotation in depth is possible. That is, human subjects can immediately exploit viewpoint-invariant information without prior familiarity with the objects.

## Acknowledgements

We thank G.E. Legge, J.T. Todd and an anonymous reviewer for helpful comments. Supported by ARO DAAH04-94-G-0065 and ONR N0014-95-1-1108.

## Appendix A. Mean correct reaction times and (% errors) for the within subjects group (rotation a) of Experiment I for all trials for the set of six objects with equal NAP and MP salience when unrotated

| Condition | Unrotated | Rotated A |
|---|---|---|
| Same | 675 (11.7) | 737 (11.4) |
| NAP difference | 725 (20.0) | 761 (22.9) |
| MP difference | 728 (13.3) | 881 (65.4) |

The unrotated condition is only for the 0° orientation. The six objects were calibrated to have equally detectable MP and NAP differences at both 0 and 57° for the unrotated trials for the between subject conditions shown in Appendix C.

## Appendix B. Mean correct reaction times and (% errors) for the within subjects group (rotation A) of Experiment I for the full set of 12 stimuli for unrotated and rotated trials

| Condition | Unrotated | Rotated A |
|---|---|---|
| Same | 750 (17.5) | 774 (17.3) |
| NAP difference | 778 (20.8) | 763 (25.0) |
| MP difference | 771 (20.8) | 909 (70.0) |

The unrotated condition is only for the 0° orientation. The six objects were calibrated to have equally detectable MP and NAP differences at both 0 and 57° for the unrotated trials for the between subject conditions shown in Appendix C.

## Appendix C. Mean correct reaction times and (% errors) for the between subjects groups of Experiment I for the all trials with the set of six objects with equal NAP and MP salience when unrotated

| Condition | Unrotated | Rotated B |
|---|---|---|
| Same | 766 (20.0) | 809 (13.2) |
| NAP difference | 844 (26.6) | 809 (24.0) |
| MP difference | 858 (27.2) | 942 (60.7) |

The unrotated data are from separate groups run at 0 and 57° orientations. The rotated B group constituted a third group.

## Appendix D. Mean correct reaction times and (% errors) for the between subjects groups of Experiment I for unrotated and rotated trials for the full set of 12 stimuli

| Condition | Unrotated | Rotated B |
|---|---|---|
| Same | 797 (20.0) | 837 (18.5) |
| NAP difference | 774 (20.0) | 829 (20.0) |
| MP difference | 774 (20.0) | 962 (64.0) |

The unrotated data are from separate groups run at 0 and 57° orientations. The rotated B group constituted a third group.

## Appendix E. Mean correct reaction times and (% errors) for the full set of 12 stimuli for Experiment II (mixed blocks) for unrotated and rotated trials

| Condition | Unrotated | Rotated |
|---|---|---|
| Same | 1134 (12.2) | 1329 (24.4) |
| NAP difference | 1155 (28.6) | 1155 (25.9) |
| MP difference | 1286 (59.5) | 1388 (55.3) |

## References

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

Biederman, I. (1988). Aspects and extensions of a theory of human image understanding. In Z. Pylyshyn, *Computational processes in human vision* (pp. 370–428). New York: Ablex.

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

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

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

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

Cohen, D., & Kubovy, M. (1993)). Mental rotation, mental representation, and flat slopes. *Cognitive Psychology*, *25*, 351–382.

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.

Edelman, S. (1995). Representation, similarity, and the chorus or 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.

Egeth, H. (1966). Parallel versus serial processes in multidimensional discrimination. *Perception and Psychophysics*, *5*, 341–346.

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

Hochberg, J., & Gellman, L. (1977). The effect of landmark features on mental rotation times. *Memory and Cognition*, *5*, 23–26.

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). Recognising novel views of three-dimensional objects. *Canadian Journal of Psychology*, *46*, 170–190.

Intraub, H. (1984). Conceptual masking: the effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology*: *Learning*, *Memory*, *and Cognition*, *10*, 115–125.

Jacobs, D. W. (1989)). Space efficient 3D model indexing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA: IEEE Computer Society, 439–444.

Jacobs, D. W. (1997). *What makes viewpoint invariant properties perceptually salient?* Princeton, NJ: NEC Research Institute Unpubl. manuscr.

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

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

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

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

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

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

Stankiewicz, B. J., Hummel, J. E., & Cooper, E. E. (1998). The role of attention in priming for left-right reflections of object images: evidence for a dual representation of object shape. *Journal of Experimental Psychology*: *Human Perception and Performance*, *24*, 732–744.

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., & Volker, B. (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.

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

Vetter, T., & Poggio, T. (1994)). Symmetric 3D objects are an easy case for 2D object recognition. *Spatial Vision*, *8*, 443–453.