Inferior Temporal Neurons Show Greater Sensitivity to Nonaccidental than to Metric Shape Differences

Rufin Vogels[1], Irving Biederman[2], Moshe Bar[2,4], and Andras Lorincz[3]

Abstract

■ It has long been known that macaque inferior temporal (IT) neurons tend to fire more strongly to some shapes than to others, and that different IT neurons can show markedly different shape preferences. Beyond the discovery that these preferences can be elicited by features of moderate complexity, no general principle of (nonface) object recognition had emerged by which this enormous variation in selectivity could be understood. Psychophysical, as well as computational work, suggests that one such principle is the difference between viewpoint-invariant, nonaccidental (NAP) and view-dependent, metric shape properties (MPs). We measured the responses of single IT neurons to objects differing in either a NAP (namely, a change in a geon) or an MP of a single part, shown at two orientations in depth. The cells were more sensitive to changes in NAPs than in MPs, even though the image variation (as assessed by wavelet-like measures) produced by the former were smaller than the latter. The magnitude of the response modulation from the rotation itself was, on average, similar to that produced by the NAP differences, although the image changes from the rotation were much greater than that produced by NAP differences. Multidimensional scaling of the neural responses indicated a NAP/MP dimension, independent of an orientation dimension. The present results thus demonstrate that a significant portion of the neural code of IT cells represents differences in NAPs rather than MPs. This code may enable immediate recognition of novel objects at new views. ■

INTRODUCTION

Humans often show little difficulty in recognizing objects from arbitrary viewpoints. Indeed, several experiments have demonstrated that for certain sets of novel objects, slight to modest effects of object rotation in depth are manifested, either by humans (Biederman & Bar, 1999; Biederman & Gerhardstein, 1993) or monkeys (Logothetis, Pauls, Bülthoff, & Poggio, 1994, for stimuli with distinctive parts). One theoretical account of this remarkable competence proposes that the observer exploits nonaccidental properties (NAPs) that are relatively invariant over rotations in depth (Biederman, 1987). A NAP is an image property, such as the linearity of a contour or the cotermination of a pair of contours, that is unaffected by rotation in depth, as long the surfaces manifesting that property are still present in the image (Lowe, 1985).[1] NAPs can be distinguished from metric properties (MPs), such as the aspect ratio of a part or the degree of curvature of a contour, which do vary continuously with rotation in depth. Indeed, the critical feature of the demonstrations showing immediate viewpoint invariance was the availability of NAPs distinguishing one object from another.

Biederman and Bar (1999) tested whether human subjects show less orientation dependency when objects to be discriminated differed in a NAP than when they differed in an MP in a generalized cylinder characterization of a single part, i.e., a geon change.[2] Their subjects viewed a sequence of two 2-part objects, each followed by a mask, and had to judge whether the objects were identical or not, ignoring differences in orientation in depth. In an object with a curved-axis cylinder on top of a brick, for example, the MP change would be a change in the angle of attachment of the cylinder to the brick whereas the NAP change would be a change in the axis of the cylinder from curved to straight.

In order to be able to compare the sensitivity for NAP versus MP changes, Biederman and Bar calibrated the magnitudes of the MP and NAP changes to be equally detectable at the same orientation-in-depth, by reaction times and error rates. (As discussed later, the MP and NAP changes were also equated in terms of image differences.) In the main experiment, subjects had to judge whether a brief sequential presentation of two gray-level images of novel two-part objects at different orientations depicted the same or different objects. Each object sequence was viewed only once. On each trial, subjects could not predict whether a change would occur and, if so, whether it would be of an MP or a NAP (viz., a geon), and which of the two

[1] KU Leuven, Belgium, [2] University of Southern California, [3] Eotvos Lorand University, Budapest, Hungary, [4] Massachusetts General Hospital

parts would undergo that change. 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 in detecting MP differences. Thus, in agreement with the privileged role of NAP differences in object recognition, novel objects differing in a NAP and presented under different orientations-in-depth were discriminated much faster and with far greater accuracy than objects differing in an MP. Although not all previous studies have evidenced similarly miniscule rotation costs when distinguishing NAPs were available (e.g., Hayward & Tarr, 1997), all have shown substantial facilitation in recognition/discrimination when distinguishing NAPs were present compared to when only metric variation was present (e.g., Tarr, Bülthoff, Zabinski, & Blanz, 1997). In the Tarr et al. (1997) study, e.g., adding a single distinguishing geon to each of a set of bent paper clip objects increased *d*s for their same–different matching from approximately 0.8 to 3.3 at a rotation angle of 60°.

Our primary interest in the present investigation was to determine the extent to which the relatively greater salience of NAP over MP differences in object discrimination is reflected by the tuning of cells in the inferior temporal (IT) cortex of the macaque, an area known to be involved in object recognition (Logothetis & Sheinberg, 1996; Ungerleider & Mishkin, 1982). The recordings were obtained in anterior IT (area TE; Figure 1) of awake, behaving monkeys performing a fixation task.

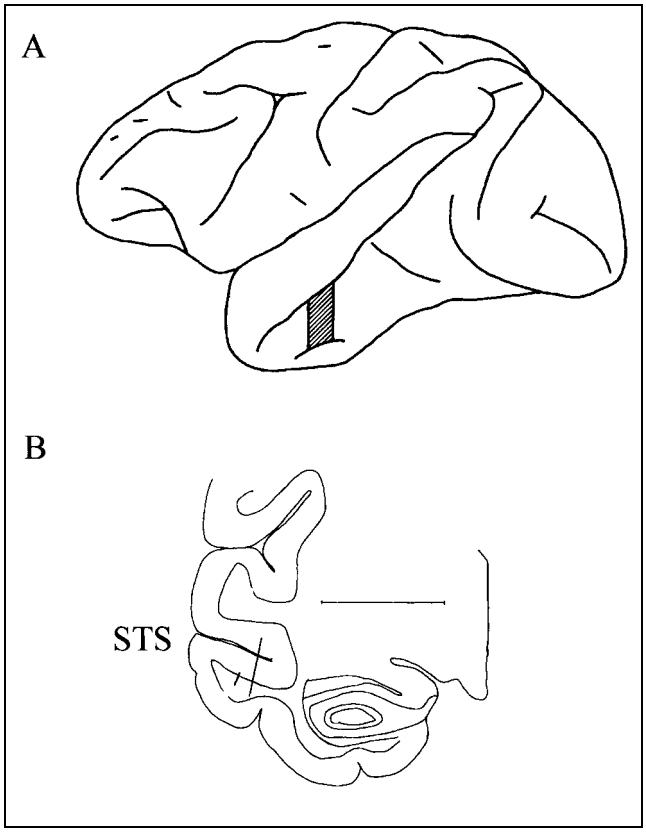

Figure 1. Recording site in IT cortex. (A) Reconstructed recording site (hatched) shown on a lateral view of a standard rhesus monkey brain. (B) Drawing of a 60-μm-thick coronal section of the temporal lobe of one of the three animals showing electrode tracks. Horizontal calibration bar: 1 cm. STS = superior temporal sulcus.

The stimuli used in the present investigation were those used in the human psychophysical study of Biederman and Bar (1999). These stimuli had the advantage of being a calibrated set of objects in terms of the MP and NAP differences. Furthermore, the objects depicted in these images were composed of only two parts, approximately matching the pattern complexity preferred by many IT cells (Tanaka, Saito, Fukuda, & Moriya, 1991). A novel feature of the present investigation was that we used several measures of image similarity to assess whether greater neuronal sensitivity to NAP than to MP changes was not simply due to a difference in the magnitude of the corresponding physical changes in the images, as assessed by measures of wavelet and pixel differences. Because changing a geon in many cases produced more local feature changes, for example, in the number of vertices (though at much smaller scales), than were produced by a change in an MP, it was necessary to evaluate the effect of the number of feature changes, per se, on the magnitude of neuronal modulation. To anticipate the results, the greater modulation of NAP changes compared to MP changes could not be attributed to image differences (as scaled by wavelets or pixels) or the number of feature changes. Thus, the single large metric change produced less neural response modulation than an equated change of a geon (or of several relatively small nonaccidental features).

RESULTS

IT neurons were tested with object images presented centrally during fixation. The images were derived from 13 ''original'' objects (the 12 experimental objects of Biederman and Bar and their practice object) and grouped accordingly in 13 object families. Each object family consisted of six images (see Figure 2 for two examples): an ''unrotated'' view of the original, an unrotated view of a NAP-changed version of the original, the same view of an MP-changed version of the original, and their three rotated versions. Responsive neurons were sought by presenting each of the 13 images of the unrotated originals. Neurons (N = 130) that were responsive to at least one of these original objects were tested further by presenting the six images of two object families (those of the two original objects that elicited the largest responses). Two examples of such neurons are shown in Figure 2. The responses of both neurons were modulated more strongly by a NAP than by an MP change at both orientations-in-depth. Furthermore, the difference in response between the unrotated original object and the rotated original object was similar to the

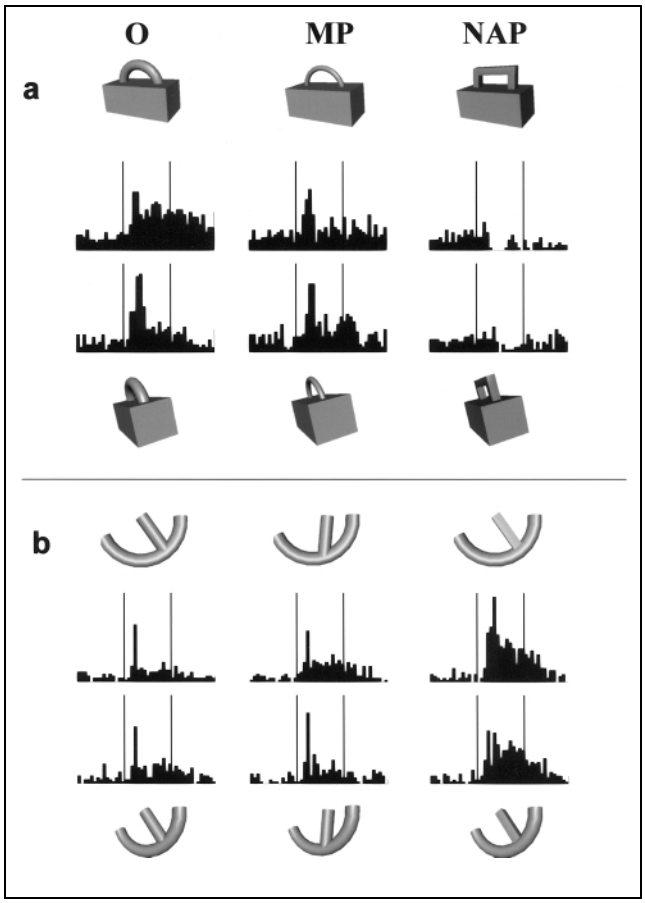

Figure 2. Responses of two IT neurons to changes in nonaccidental (NAP) and metric shape properties (MP) under rotation. (a, b) Poststimulus time histograms of responses of two IT neurons to the original (O), NAP-changed, and MP-changed images and their rotated versions. The stimuli are shown on top (unrotated conditions) and below (rotated) the histograms. The two vertical lines indicate stimulus appearance. The heights of the vertical lines correspond to 220 and 115 spikes/sec in (a) and (b) respectively. For both (a) and (b), the neuron responded similarly to MP and O images, at both orientations. For (a), the NAP image resulted in reduced firing compared to MP and O; for (b), the NAP image produced increased firing compared to MP and O.

difference in response between the unrotated original and rotated MP versions, but differed greatly from the difference in response between the unrotated original and rotated NAP versions. The latter is exactly what one would predict when neuronal response modulations underlie the greater behavioral detection of NAP versus MP changes of rotated objects as reported by Biederman and Bar for the same images.

Since each of the 130 responsive IT neurons was tested with two object families, one could, in principle, compare the effect of the NAP versus MP changes in $2 \times 130 = 260$ cases. There was a significant response to least to one of the six images of an object family in 241 cases, yielding 241 cases in which one can effectively compare the effect of NAP and MP changes. (For 19 neurons the response to the second object family was not significant.) In Biederman and Bar's behavioral study, the subjects had to decide whether or not each of the three possible rotated versions (original, MP change, and NAP change) of an object family depicted the same object as the unrotated original image. Only those neurons that show a difference in response among the three rotated versions can provide information regarding the identity of the object under rotation. Following this logic, a population analysis of the present neuronal data was done on those 110 cases in which there was a statistically significant difference in response among the three rotated images. The 110 cases consisted of responses of 80 neurons, 50 of which contributed one case (or object family) to the data and 30 of which contributed two cases. Since the three different animals showed a similar pattern of results and to increase the power of the statistical testing, the data from the three animals were pooled.

Response modulations were expressed in two ways. First, for each case we computed the absolute differences in net response between different images, for example, between rotated NAP versions and unrotated originals, and these differences were averaged over cases (Figure 3A). Second, these absolute differences were expressed as a percent of the response to the unrotated original versions, and subsequently averaged over cases (Figure 3B). Note that there was a response to the unrotated original in each case. The pattern of results was similar in the two analyses. Rotated NAP versions resulted in greater mean response changes (Figure 3A) than rotated original objects, with the response to the unrotated originals as baseline (Scheffe post hoc test: $p < .05$ in both analysis), whereas the response changes for the rotated MP versions did not differ significantly from the response changes due to the rotation of the original objects (Scheffe post hoc test; *ns*). The percent response change (Figure 3B) for the rotated NAP objects was significantly greater than the percent response change of the rotated MP objects [$t(109) = 1.73$; $p < .05$; one-tailed paired t test]. The same trend was present for the absolute mean response changes, but failed to reach statistical significance [$t(109) = 1.46$; $p < .07$].

Figures 3A and B also show the response modulations for rotated NAP and MP changes with the rotated originals as a baseline (last two columns in panels A and B of Figure 3). For the rotated views, changing a NAP produced significantly larger response modulation than changing an MP (paired t test, $p < .001$ in both analyses). The greater NAP versus MP differences in the percent change measure is at least partially due to the low responsiveness of some of these neurons to the ''rotated original'' (giving a division by a small number and thus large modulation). If only those cases in which the response to the rotated original was at least 6.7 spikes/sec (i.e., two spikes in the 300-msec interval) are included, the mean modulation for the rotated NAP-changed objects was 64% but only 38% for the rotated MP-changed objects.

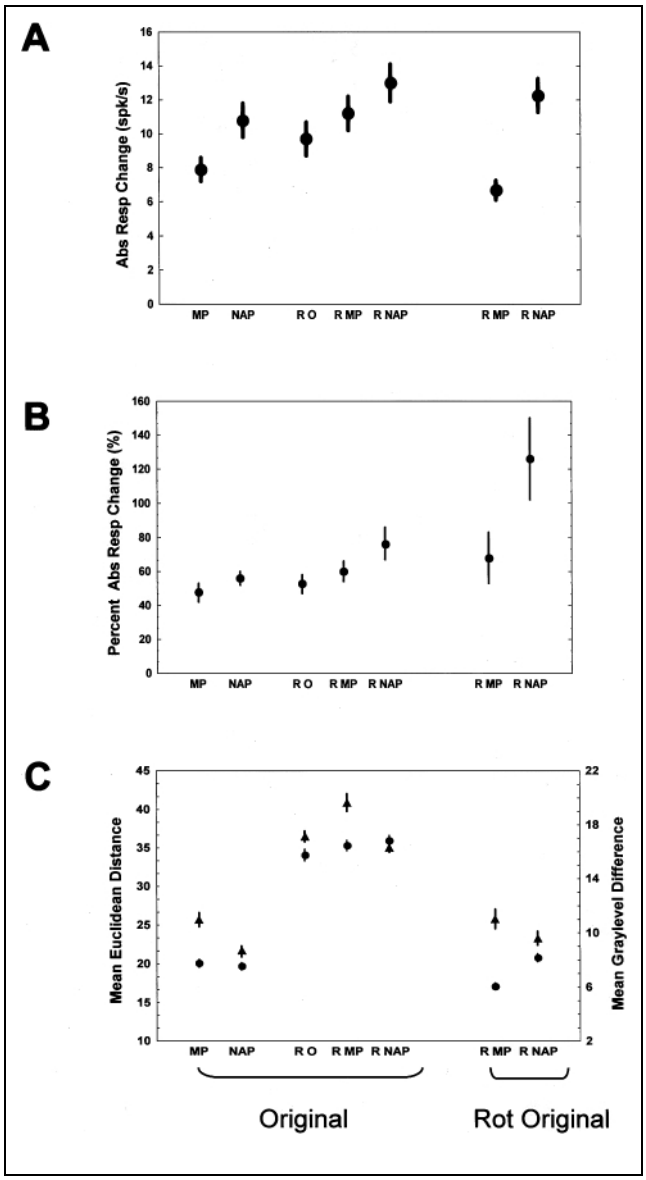

Figure 3. Mean neural response modulations to NAP and MP image changes. (A) Mean absolute differences in neural response (n = 110) between pairs of images of the same object family. The five left data points show the mean absolute difference in response to the unrotated original object and the image versions indicated on the abscissa, while the two right data points show the mean absolute difference in response between the rotated original and the two other rotated image versions. Thin bars indicate standard errors. (B) Mean percent differences in neural responses (n = 110) for the same image pairs as (A). The absolute differences were expressed as a percent of either the response to the unrotated original (5 left data points) or the response to the rotated original (2 right data points). (C) Mean Euclidean distances in wavelet space (triangles) and mean position-corrected gray-level difference per pixel (circles), indicating the magnitude of the physical differences between images. Same conventions as for (A, B). MP = MP-changed object; NAP = NAP-changed image; R = rotated condition; O = original.

For the unrotated views (first two columns, Figure 3A and B), the response modulations were also consistently larger for NAP than for MP changes, but this difference reached statistical significance only for the absolute response differences, $t(109) = 2.8$; $p < .005$, and fell short of significance when the modulation was expressed as a percent of the response to the original objects, $t(109) = 1.4$; $p < .08$.

Do these neuronal response modulations merely reflect physical similarities between the different images? No. Physical image similarity was assessed with two kinds of measures (described in more detail in the Methods section): (a) the similarity of a pair of images was expressed as the Euclidean distance between the images in ''wavelet space,'' and (b) the mean absolute difference in gray level per pixel between two images (i.e., Hamming distance), corrected for possible position shifts. These measures capture all image variations, including those that are produced by differences in surface illumination, as well as orientation and shape (i.e., MP and NAP) changes. Figure 3C shows the average wavelet-based and gray-level-based image similarities for the same image pairs of which the neuronal modulations are shown in Figure 3A and B. Both physical similarity measures produced similar image similarity rankings, with one exception. For the wavelet measure, the rotated MP-changed images were slightly more dissimilar from the originals than the rotated NAP-changed images, which is the opposite of what would be expected from the neuronal response modulations. For the position-corrected pixel differences, however, there was slightly greater similarity of the MP-changed stimuli compared to the NAP-changed stimuli.

As a further test of whether image similarity could account for the greater modulation of NAP compared to MP changes, we assigned each object family to one of two groups, based on the position-corrected, pixel gray-level differences. In one group, the difference in physical similarity between the rotated MP-changed images and the unrotated originals was larger than the physical difference between the rotated NAP-changed images and the unrotated original (six families; n = 55 cases), and in the other group the opposite was the case (seven families; n = 55 cases). Neuronal response modulations were computed for the two groups separately and these are shown in Figure 4A. The interaction between the grouping variable and the response modulation was not statistically significant, $F(2,216) < 1$. This was also the case when the neuronal modulations were computed as absolute response differences (not shown). Thus, the neuronal modulations produced by the MP and NAP variations do not merely reflect physical image similarities. A clear case of this can be seen in a comparison of the rotated originals and rotated NAP-changed objects for one group (diamonds in Figure 4B). Whereas the position-corrected image similarities compared to the original objects were virtually identical for the two conditions, the neuronal modulations for the rotated

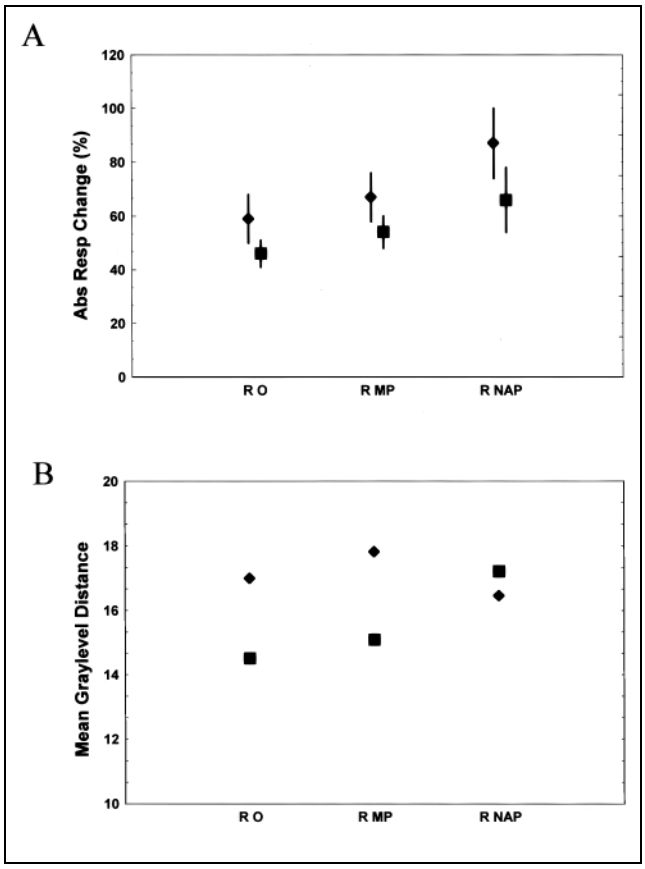

Figure 4. Comparison of neuronal response modulations and physical image similarities. (A) Mean percent absolute response differences between originals and rotated originals, rotated MP images, and rotated NAP images for two groups of object families, those in which the position-corrected gray-level image distances (against the originals), shown in (B), were larger for rotated NAP changes compared to the rotated MP changes (squares) and those in which the reverse was true (diamonds).

NAP changes were significantly larger than for rotated originals of that group (diamonds in Figure 4A).

The previously described analyses of the response modulations were performed on a selected sample of neurons, namely those responding differentially among the rotated versions. However, it is also important to ascertain how the other neurons respond to these object changes and whether the NAP/MP difference really is a critical factor affecting the responses of the whole population of responsive IT neurons. Furthermore, in the above analysis, only 7 of the 15 possible comparisons between the six images of an object family were analyzed. The additional 8 comparisons would produce a more complete picture of how these image changes affect IT neurons and, thus, how IT neurons can represent differences between these images. The latter questions were addressed by applying multidimensional scaling (MDS) to the neural responses. MDS represents the variation in firing rates across the different stimuli as distances in a low-dimensional space, allowing one to detect the principal underlying dimensions of the representation. The MDS was performed on Euclidean distances computed using all cases (n =215) in which the maximal response was at least 10 spikes/sec (mean net response = 33 spikes/sec). Two dimensions of the MDS were sufficient and necessary to represent the neural distances between the six image types (Figure 5). We suggest that Dimension 1 is a NAP change versus MP change/original dimension and that Dimension 2 corresponds to the viewpoint change. Note the tight clustering of the MP and original image versions compared to the much larger distances between the original and the NAP-changed images. In fact, for this larger, unselected sample of cases, the absolute response changes against the original images were significantly larger for NAP-changed than for MP-changed objects, for both unrotated, $t(214) = 2.9$, $p < .002$, and rotated images, $t(214) = 4.5$, $p < .00001$. As for the smaller sample of neurons (see above), the mean change in response between the unrotated original and the rotated NAP versions differed significantly (Scheffe post hoc test: $p < .05$) from the mean change in response between the unrotated and rotated original objects, but the response changes for the rotated MP versus unrotated original objects were not close to being significantly different from the response changes due to the rotation of the original objects ($p < .37$). These results of the MDS analysis show that when equated for image changes, changing a NAP of a generalized cylinder characterization of an object part (or NAP features of

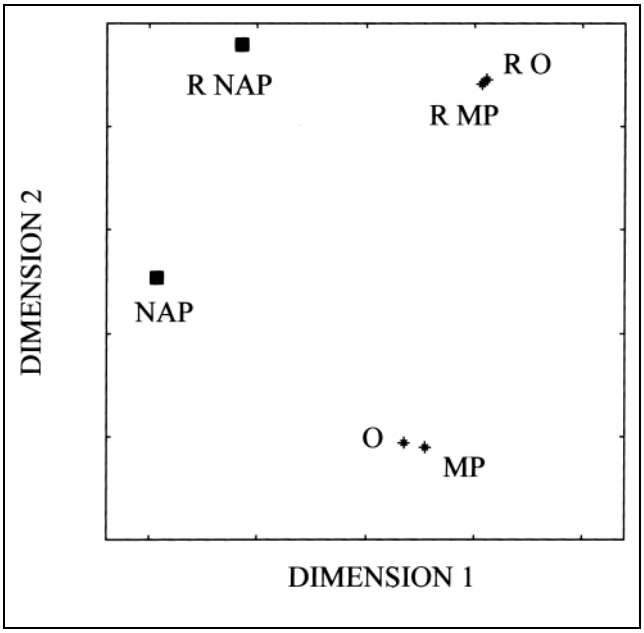

Figure 5. Two-dimensional representation of the six image types as obtained by MDS of the single cell response modulations (n = 215). The axes have the same scale. Squares: NAP-changed objects; stars: MP-changed and original objects. For other conventions see legend, Figure 3. Dimension 1 can be interpreted as NAP versus an MP/O contrast; Dimension 2 as a rotated–unrotated contrast.

that part) has a much stronger effect on the neural representation of that object than changing an MP.

DISCUSSION

In the present study we compared the effect on the response of IT neurons of NAP and MP changes. The responses of IT neurons were affected more strongly by NAP than by MP changes, and this was consistently found within as well as across views of the same object. Before discussing the implications for shape coding in macaque IT and for object representations in general, we will address methodological issues related to the present study.

Image-Based Metrics for Scaling Image Similarity

An important feature of our design was the use of image-based metrics (wavelet and gray-level analyses) for scaling qualitatively different shape differences, so that the sensitivity of neurons could be compared for the scaled shape differences. Only by doing this can one assess whether neural response modulations are determined by the stated shape differences rather than merely reflecting physical image similarity. The question of course is which similarity metric to use. We used two metrics that did not make any commitment to differential sensitivities for higher-order features. In fact, neither of the two metrics favor NAP or MP variations; changes in the image by either source of variation are treated equally by these metrics. The gray-level analysis compares image similarities as are present in the retinal input (corrected for position differences). Thus, any differential sensitivity to NAP versus MP changes has to originate within the visual system and cannot be an artifact of retinal image dissimilarities. It should be noted that these scaling measures capture variations in shading due to the NAP, MP, and/or rotation changes, implying that the greater IT sensitivity to NAPs than to MPs cannot be a consequence of differences in surface characteristics. The wavelet metric measures image distances using a sparse representation of an overcomplete basis set of Gabor-like functions.[3] Insofar as aspects of the selectivity of earlier stages in the ventral pathway (e.g., V1, V2, and V4) can be modeled by wavelets, the observed lack of correlation between wavelet-based image similarities and neuronal response modulations would suggest that the dissociation of wavelet similarity and neural activity arises in later rather than in the earlier stages. Our results are thus consistent with that of Kobatake and Tanaka (1994) who showed a gradient of declining sensitivity to wavelet components of stimuli as their recordings proceeded rostrally from V2 to TE.

Nonetheless, the lack of correlation between the magnitude of neural modulation for NAP and MP differences and the wavelet similarity measure in the present study was, most likely, at least partly due to the very small range of stimulus variation that was necessitated by the original calibration of NAP and MP differences. In general, we would expect greater neuronal modulation from images of greater dissimilarity (e.g., as in the case of rotation versus NAP changes).

Given the ''arbitrary'' and often subtle nature of particular NAP and MP changes in the present study, our finding of statistically significant differences in response modulation for NAP and MP changes, especially in the large population sample (Figure 5), is dramatic, and suggests a strong, overall sensitivity bias for NAP changes compared to MP changes.

Consistent with our findings of smaller modulation to MP differences is a report that many IT neurons show a relatively weak sensitivity to variations in the aspect ratio (an MP) of an ellipse, as would be produced by rotating the ellipse in depth (Esteky & Tanaka, 1998). However, the Esteky and Tanaka investigation did not contrast NAP and MP changes as done in the present study. Further evidence for a differential sensitivity of IT neurons for qualitative versus metric changes comes from a recent study of 3-D-shape selectivity of IT neurons (Janssen, Vogels, & Orban, 1999). Changing the sign of binocular disparity so that a convex shape became concave (a qualitative NAP change) produced a much stronger response difference in IT neurons than even larger disparity changes by which the degree of shape convexity was manipulated, which is an MP change.

When designing the stimuli of the present study, care was taken to minimize geon differences between the two views of the same object. Nonetheless, relatively strong effects of rotation of the objects were present (Figures 4 and 5), in agreement with Logothetis and Pauls (1995). However, this does not contradict our claim that IT neurons are more sensitive to NAP than to MP changes, insofar as the physical differences between images of different orientations of the same object were, on average, much larger than those between two objects at the same orientation, differing only in a NAP (Figure 3C).

Number of Feature Changes

The NAP manipulation in the present experiment corresponded to a change in a geon. As noted in the Introduction, a change in a geon can produce changes in several vertices and edges, particularly when a crossection changed from curved to straight or vice versa (Biederman, 1987). NAP-changed images compared to the originals thus often differed in more features than those that distinguished MP-changed images from the originals. It is possible, then, that the greater modulation for the NAP-changed stimuli could be a consequence of the larger number of NAP feature changes rather than to the nonaccidental change in the generalized-cylinder characterization of the geon, for example, whether the axis or cross section was straight

or curved. To assess the possible effect of variation in the number of feature changes, we correlated the degree of response modulation and the number of features that changed among the different conditions. All feature changes—the presence or absence of vertices, edges, and surfaces, a change in a vertex (e.g., from a fork to an L for), edge (e.g., straight to curved), the parallelism (vs. nonparallelism) of pairs of edges, and brightness feature changes (such as the presence vs. absence of a ''hot spot'')—were counted by two judges, whatever the size of the features (details and reliability measures in method section). The Pearson correlation between the number of feature changes and the response modulations for the unrotated NAP–unrotated original comparison were computed using all cases for which the response to the original or to the NAP-changed version was equal or larger than 10 spikes/sec. The latter prevents the inclusion of weakly responsive neurons. The correlation between the number of feature changes and the degree of response modulation was negligible and not statistically significant (percent response change: $R = -.08$ ($n = 190$; *ns*); absolute response change: $R = .03$ ($n = 190$; *ns*). Thus, the larger modulations for NAP versus MP changes was likely not due to a difference in the number of features that were changed in the NAP compared to the MP, but to the nature of the feature change (NAP vs. MP) that characterized a geon attribute.

Undoubtedly, the lack of an effect of the number of feature differences is in large part attributable to the very small scale—both absolute and relative to the size of the object—of most of the feature changes, particularly those of the vertices (which accounted for most of the feature differences). The wavelet and gray-level measures of image similarity described above take into account all feature changes. Because only a single (metric) feature typically changed for the MP-changed stimuli, the NAP features would necessarily have been of small scale to produce an equivalent, much less small, difference in image similarity. We acknowledge that a more definitive assessment of the effects of the number of features and their scale of variation awaits direct systematic tests of these variables. A slightly weaker conclusion of the present experiment could be that, when equated for the magnitude of stimulus change, a single, relatively large metric change produces less modulation than: (a) a change of a geon, and (b) several relatively small nonaccidental feature changes.

Correlation of Single Cell Responses and Human Psychophysical Performance

The larger neuronal response modulations for NAP versus MP changes is consistent with theory (Biederman, 1987) and human psychophysics (Biederman & Bar, 1999). It is extremely likely that monkeys show a corresponding larger behavioral sensitivity for NAP versus MP changes under rotation. Indeed, Logothetis et al. (1994) reported that for novel objects that have distinguishing parts, for example, a spaceship, monkeys show immediate viewpoint invariance, as do humans.

Humans and, presumably, monkeys are much more sensitive for NAP changes than for equivalent MP changes when comparing objects at different or at the same orientations-in-depth. In the Biederman and Bar (1999) study, the equivalence in detecting MP and NAP changes at the same orientation held only when the objects, throughout a block of trials, were always presented at the same orientation. When same versus different orientation was varied within a block in random-appearing fashion (Experiment 2 of Biederman & Bar), the detection of MP differences fell to chance levels, whereas RTs and error rates for the detection of NAP differences were only slightly increased. Our finding of a greater IT response modulation for NAP versus MP changes of objects at the same or different orientations-in-depth fits these psychophysical data from the mixed block. The ability to detect subtle MP differences of novel objects in brief presentations may require conditions in which orientation is held constant.

The strong effects of rotation render a direct link of the neuronal response modulations and behavioral decisions difficult. The average neural response difference due to rotating an object was of similar magnitude to changing a geon (Figures 3A and B). This implies that not only objects differing in NAPs, but also images of the same objects at different orientations-in-depth produce highly different activation of neuronal populations in IT. How can an observer know whether different population activities correspond to different objects or, instead, to different views of the same object? This is a general problem: It also arises when attempting to link, for example, V1 responses and orientation discrimination. Subjects can judge extremely accurately the orientation of a grating that varies in spatial frequency from trial to trial (Vogels, Eeckhout, & Orban, 1988; Burbeck & Regan, 1983). Since single V1 neurons are tuned for orientation and for spatial frequency, response modulations can be due to either a difference in orientation or spatial frequency. One way to solve this problem is to pool across neurons of the same orientation preference (or across neurons of the same spatial frequency preference if one needs to compute stimulus spatial frequency). This, however, requires ''labeling'' of the neuronal stimulus preferences. One can envisage a similar strategy in IT, i.e., a pooling of activity across cells tuned to different views of the same object part. At the level of V1, the orientation preference of a neuron can be labeled by virtue of the columnar organization of orientation. Similarly, IT neurons could be organized in columns according to NAP differences and cells within the same column could show different tunings for large MP changes of the same basic geon. Fujita, Tanaka, Ito, and Cheng (1992) have reported evidence for a columnar

organization of moderately complex features in IT, but it is not clear how well this would fit the above postulated organizational scheme.

The presence of ''immediate'' view-invariant behavior (Biederman & Bar, 1999; Logothetis et al., 1994) implies that such labeling of neurons responsive to different views of the same, novel object must exist before exposure to the particular object. This can be done by having labeled units that have a relatively broad tuning along some shape dimensions but a sharper one along other NAP dimensions. The broad tuning of these neurons allows them to respond to novel parts but, because of their narrow tuning along other NAP dimensions, they will, as a population, produce a NAP-selective signal. Immediate view invariance will be facilitated when the units are much less sensitive to MP changes. It is possible that when the animals are matching objects over different orientations, the neuronal response modulation from rotation (large MP changes) itself is strongly reduced while those for NAP changes are enhanced. Alternatively, and perhaps more plausibly, those neurons showing minimal effects of rotation, but strong NAP modulation, are selected when performing object matching. Given the ubiquity of object constancy, it may well be that it is the rotation-invariant neurons that mediate our default state of object awareness.

METHODS

Stimuli

The gray-level images were the same as those in the human psychophysical experiment (Biederman & Bar, 1999). The stimulus set consisted of 13 object families, two of which are illustrated in Figure 2. Each object family was derived from 1 of 13 original, two-part, objects. One of the parts of the original object was changed in either an MP or a NAP, that is, a geon, and these three objects (original, NAP, and MP versions) were rendered at two orientations-in-depth, yielding six images per object family. One of the two viewpoints was arbitrarily chosen as ''unrotated'' and the other as ''rotated.'' The objects were rotated around the vertical an average of 57° (range 20° to 120°). For details of the calibration of the NAP and MP differences, see Biederman and Bar (1999). The magnitude of the metric differences was constrained to not obviously alter the relative relations among the object's parts, for example, so that a small part became a larger part. To achieve the calibration, the NAP changes had to be subtle and correspond to differences among highly similar subordinate members of an object class. The images (size ± 5°; luminance gamma corrected) were shown at the center of a Phillips 21-in. display.

Image-Similarity Measures

Differences between the images were computed pairwise within each object family using two measures. One measure was based on a wavelet analysis of each of the images (Daubechies, 1988). The Euclidean distance between the wavelet coefficients for the different images, after lossy compression (Gibson, Berger, Lookabaugh, Lindbergh, & Baker, 1998) was then taken as a measure of physical similarity. The other measure consisted of the absolute difference in luminance, calculated pixelwise and then averaged over all pixels of the image (Adini, Moses, & Ullman, 1997) with a correction for position shifts (described below). Note that since we used the wavelet space expansion for lossy compression, the wavelet-based distances are not identical to Euclidean distances between pixelwise gray-level values of image pairs.

The above physical similarity measures are affected not only by feature changes but also by the relative position of the objects in the images. This can be illustrated by the following example. Consider three images, one with a one-pixel-wide vertical bar, a second one with the same bar shifted horizontally by one pixel, and a third one consisting of a horizontal bar. The pixelwise gray-level distance between the first image and the position-shifted image will be larger than between the orientation-changed images. However, an orientation-tuned neuron responding in a position-invariant way, as many neurons in IT do for these small position differences, will show a much stronger modulation for the orientation change than for the position change. Although it is unlikely that the position differences between the NAP-changed and original were systematically larger than between the MP-changed and original objects, we nevertheless computed gray-level similarity for pairs of images, corrected for position differences.

To compute image similarities corrected for position differences between objects, we computed the pixelwise (absolute) gray-level difference for different relative positions of the object. One object was systematically shifted by one pixel (range 0–80 pixels) in the vertical and/or horizontal directions and for each shift the gray-level difference was computed. The minimum of gray-level differences of the set of (2×80 [H] $\times$ 2×80 [V]) = 25,600 relative positions was then taken as the position-corrected gray-level image similarity.

The data shown in Figure 3C are weighted averages of the image distances. The distance of a particular image pair was weighted according to its frequency in the neuronal database.

Task and Single Cell Recording Procedures

Trials started with the onset of a small fixation target at the display's center on which the monkey was required to fixate (eye position measured with scleral search coil technique). After a fixation period of 700 msec, the fixation target was replaced by the stimulus for 300 msec, followed by presentation of the fixation target

for another 100 msec. If the monkey's gaze remained within a 1.5° fixation window until the end of the trial, he was rewarded with a drop of apple juice.

Single IT neurons were recorded with tungsten electrodes, lowered in a guiding tube, using a vertical, dorsal approach. The position of the recording chamber was determined preoperatively with structural MRI. Verification of recording positions was done by superimposing the MRI images and images of the skull obtained with a spiral CT scan (with guiding tube in situ). Based on the recording depth with respect to the bone and gray/white matter transitions, the neurons are from the lower bank of the superior temporal sulcus and, mainly, from the lateral convexity of area TE (Figure 1A). Histological confirmation of the recording sites of two of the three animals is available (Figure 1B). All surgical procedures and animal care was in accordance with the guidelines of NIH and of the KU Leuven Medical School.

Design and Data Analysis

After isolating a neuron responsive to at least one of the 13 original objects, 12 images were presented interleaved (the order of presentation of the 12 images was randomized in blocks of 12 trials each; at least 10 blocks of 12 trials (i.e., at least 10 trials/image) were run). The 12 images consisted of the unrotated and the rotated original, NAP- and MP-changed versions of two objects. The two object families that were chosen for testing a given cell were those whose original objects (of the 13) elicited the most (and second most) activity for that cell.

For each trial, spikes were counted in windows of 300-msec duration. The baseline activity, obtained in a window preceding stimulus onset, was subtracted from the stimulus-induced activity, measured in a window starting 50 msec after stimulus onset. Statistical significance of responses was assessed by ANOVA, which compared the spike counts in the two windows. Other ANOVAs, a priori, and post hoc comparisons were performed on the net responses to test for stimulus selectivity and differences between conditions.

For the MDS, neural distances were computed in three ways: Manhattan (City Block) distances, Euclidean distances and Pearson correlation coefficients. In this implementation of nonmetric MDS (Statistica for Windows, Statsoft, Tulsa, OK), the starting configuration is based on principal component analysis and is thus unbiased. The configurations were similar for the different distance measures. The final 2-D configuration of Figure 5, which is based on Euclidean distances, fit the observed distances very well (coefficient of alienation: 0.00002; see Guttman, 1968), while a 1-D configuration provided a worse fit (alienation: 0.26161). This indicates that two dimensions were necessary and largely sufficient to represent the observed neural distances. In order to assess the reliability of the configuration, we randomly assigned each case to one of two groups, and then performed MDS on the Euclidean distances computed for each group separately. The configurations of these two independent MDSs were highly similar (correlation of "D-star" distances: $R = .98$, $n = 15$), indicating that the configuration of Figure 5 is reliable.

Judgment of the Number of Feature Differences

Two raters counted the number of feature changes between all pairs of stimuli within each stimulus family. The raters were instructed to include differences in the presence (vs. absence) of a line, a change in curvature from curve to straight, a change or presence–absence of a vertex, a change in the parallelism of lines, and changes in shading and luminosity features. The scale of the change was to be ignored except that differences that were judged to be not perceptible at a brief duration were not be counted. Test–retest reliability (Pearson R) of Judge 1 was 0.94; for Judge 2 it was 0.79. Correlation of the average of Judge 1's ratings over the stimulus comparisons with the average of those of Judge 2 was 0.79. Most of the variation in the ratings was due to differences in the judged perceptibility of very small changes.

Acknowledgments

The technical help of M. De Paep, P. Kayenbergh, G. Meulemans, G. Vanparrys, and C. von der Malsburg, the critical reading of an earlier version of this paper by Gy. Sary, P. Janssen, and A. Rossier, and the ratings by M. C. Mangini and E. A. Vessel are gratefully acknowledged. Supported by Geneeskundige Stichting Koningin Elizabeth (R.V.), GOA 95-99/06 (R.V.), BIL97/31 (R.V./A.L.), ARO DAAHO4-94-G-0065 (I.B.), ARO DAAG55-97-1-0185 (I.B.) and grant RG0035/2000-B from the Human Frontiers Science Program.

Reprints requests should be sent to: Rufin Vogels, Laboratorium voor Neuro- en Psychofysiologie, Faculteit der Geneeskunde, KU Leuven, Campus Gasthuisberg, Herestraat, B-3000 Leuven, Belgium. E-mail: rufin.vogels@med.kuleuven.ac.be.

Notes

1. Formal analyses of the basis of NAPs are presented by Jacobs (2000). He argues that there are an infinite number of NAPs of high dimensionality but these are not perceptually salient. NAPs that are psychologically salient, according to Jacobs, are those that have low dimensionality, in that they can be defined by one, two, or three primitive features, such as points or lines. The NAPs considered in the present work are all of low dimensionality. Zetzsche and Krieger (1999) and Krieger and Zetzsche (1996) demonstrate how NAPs might be obtained by nonlinear filtering and Barth, Zetzsche, and Renchler (1998) show how the outputs of such filters can provide texton-like features in texture segregation tasks. Koenderink (1984) has described the NAPs provided by smooth, solid shapes. NAPs have played a primary role in a number of computer vision models, for example, Zerroug and Nevatia (1996), Dickinson, Rosenfeld, and Pentland (1992), and Lowe (1987).
2. A generalized cylinder (or cone) is the volume swept out by translating a planar shape (the cross section) along an axis (Binford, 1981). Generalized cylinders afford a general way of describing simple volumes. Translating a round shape along a

straight axis produces a cylinder. A straight cross section, such as a rectangle, which would be nonaccidentally different from a round cross section, produces a brick. Geons can be specified, in part, as a partition of the set of simple generalized cylinders based on nonaccidental differences in the generating function, such as between a round and straight crossection. If the crossection varies in size (say, expands), a cone or a wedge is produced (with round and square cross sections, respectively) with nonparallel sides (in contrast to the parallel sides of the cylinder and brick). Similarly the axis can be straight or curved. The ends of volumes could be truncated, converge to a point, or be rounded. Geons also include a set of simple 2-D shapes, such as circle, square, triangle, rectangle, and ellipse. See Biederman (1987,1995) for a fuller treatment of geons. Changing a crossection from round to straight typically produces several nonaccidental feature differences in the lines and vertices composing the volume (Biederman, 1987), requiring the analysis presented in the Results section to rule out the possibility that the greater modulation of the NAP changes was not due to a larger number of feature changes (present at a smaller scale) in the NAP condition compared to the metric condition. Zerroug and Nevatia (1996) demonstrate a system that can extract the generalized cylinder structure of an object from a single gray-level image.

3. A representation is overcomplete if the elements of the representation outnumber the dimension of the input.

REFERENCES

Adini, Y., Moses, Y., & Ullman, S. (1997). Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, *19,* 721–732.

Barth, E., Zetzsche, C., & Renchler, I. (1998). Intrinsic two-dimensional features as textons. *Journal of the Optical Society of America A, 15,* 1723–1732.

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

Biederman, I. (1995). Visual object recognition. In S. M. Kosslyn, & D. N. Osherson (Eds.). *An invitation to cognitive science,* 2nd ed. *Visual cognition* (vol. 2, chap. 4, pp. 121–165) MIT Press.

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

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

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

Burbeck, C. A., & Regan, D. (1983). Independence of orientation and size in spatial discriminations. *Journal of the Optical Society of America, 73,* 1691–1694.

Daubechies, I. (1988). Orthonormal bases for compactly supported wavelets. *Communications Pure and Applied Mathematics, 41,* 909–996.

Dickinson, S., Pentland, A., & Rosenfeld, A. (1992). 3-D shape recovery using distributed aspect matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15,* 771–784.

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

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature, 360,* 343–346.

Gibson, J. D., Berger, T., Lookabaugh, T., Lindbergh, D., & Baker, R. L. (1998). *Digital compression for multimedia: Principles and standards.* San Francisco: Morgan Kaufman.

Guttman, L. (1968) A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrica, 33,* 469–506.

Hayward, 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.

Jacobs, D. W. (2000). What makes viewpoint invariant properties perceptually salient: A computational perspective. In: K. Boyer, & S. Sarkar (Eds.), *Perceptual organization for artificial vision systems*. Boston: Kluwer Academic Publishers.

Janssen, P., Vogels, R., & Orban, G. A. (1999). Inferior temporal neurons are selective for small differences in 3D structure. *Society for Neuroscience Abstracts, 25,* 529.

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

Koenderink, J. (1984). What does the occluding contour tell us about solid shape? *Perception, 13,* 321–330.

Krieger, G., & Zetzsche, C. (1996) Nonlinear image operators for the evaluation of local intrinsic dimensionality. *IEEE Transactions on Image Processing, 5,* 1026–1041.

Logothetis, N. K., & Pauls, J. (1995). Viewer-centered object representations in the primate. *Cerebral Cortex, 5,* 270–288.

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

Logothetis, N., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19,* 577–621.

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

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

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

Tanaka, K., Saito, H., Fukuda, Y., & Moriya, M. (1991). Coding visual images on objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology, 66,* 170–189.

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.

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

Vogels, R., Eeckhout, H., & Orban, G. A. (1988). The effect of feature uncertainty on spatial discriminations. *Perception, 17,* 565–577.

Zerroug, M., & Nevatia, R. (1996). Three-dimensional descriptions based on the analysis of the invariant and quasi-invariant properties of some curved-axis generalized cylinders. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18,* 237–266.

Zetzsche, C., & Krieger, G. (1999). Nonlinear neurons and higher-order statistics: New approaches to human vision and electronic image processing. In B. Rogowitz & T. V. Pappas (Eds.), *Human vision and electronic imaging IV, Proceedings of SPIE*, 3644 (pp. 2–33). WA: SPIE Bellingham.