Vision Research 69 (2012) 23–31

Contents lists available at SciVerse ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Object recognition using metric shape

Young-Lim Lee [a], Mats Lind [b], Ned Bingham [c], Geoffrey P. Bingham [a,*]

[a] Indiana University, Bloomington, IN, USA
[b] Uppsala University, Uppsala, Sweden
[c] Cornell University, Ithaca, NY, USA

ARTICLE INFO

Article history:
Received 28 December 2011
Received in revised form 19 July 2012
Available online 3 August 2012

Keywords:
Object recognition
Shape perception
Structure-from-motion
Stereo vision
Two Visual System theory

ABSTRACT

Most previous studies of 3D shape perception have shown a general inability to visually perceive metric shape. In line with this, studies of object recognition have shown that only qualitative differences, not quantitative or metric ones can be used effectively for object recognition. Recently, Bingham and Lind (2008) found that large perspective changes ($\geqslant 45°$) allow perception of metric shape and Lee and Bingham (2010) found that this, in turn, allowed accurate feedforward reaches-to-grasp objects varying in metric shape. We now investigated whether this information would allow accurate and effective recognition of objects that vary in respect to metric shape. Both judgment accuracies (d') and reaction times confirmed that, with the availability of visual information in large perspective changes, recognition of objects using quantitative as compared to qualitative properties was equivalent in accuracy and speed of judgments. The ability to recognize objects based on their metric shape is, therefore, a function of the availability or unavailability of requisite visual information. These issues and results are discussed in the context of the Two Visual System hypothesis of Milner and Goodale (1995, 2006).

1. Introduction

In a landmark study, Ballard, Hayhoe, and Pelz (1995) found that people use perception in preference to memory when guiding certain actions. When reproducing a pattern constructed of Lego blocks using a resource set of randomly organized blocks off to the side, participants looked first at the model pattern, then at the corresponding targeted block among the resource blocks and initiated their reach to it, then as the reach proceeded, they looked back at the model (using perception instead of memory to determine where to place the block) and then down to where the reproduction was being constructed as the hand transported the selected resource block to be placed in the reproduced pattern.

Participants performed the reach-to-grasp the resource block under feed forward control, not with continuous online visual guidance. This is undoubtedly common in the performance of repetitive pick-and-place tasks. Imagine a related task, for instance, sorting newly dug fossil bones into boxes representing different categories (that is, shapes) of fossilized bones. Each grasp of these complexly shaped 3D objects would be taking place as the sorter looked ahead to the box into which the bone would be placed. This sorting task would entail two sub-tasks, each requiring perception of the metric shape of the fossil objects. First, each feed forward reach-to-grasp would require perception of metric shape to guide accurate grasping. Lee and Bingham (2010) showed that this was possible. Second, object recognition entailed by the sorting task would require perception of metric shape to discriminate, for instance, the fossilized claw of the smaller *Megalonyx leptostomus* from that of the larger *Megalonyx jeffersonii* (each are species of ground sloth). In the current study, we investigated the latter possibility, that is, object recognition using perception of metric shape.

Perotti et al. (1998) had observers judge either qualitative or quantitative variations in shape. Qualitative variations in smooth surface shape (cylindrical, ellipsoidal, saddle, etc.) were measured by the "shape index" (Koenderink, 1990). Quantitative variations in shape corresponding to the amount of surface curvature were measured by the "curvedness." Perotti et al. found that the shape index was judged accurately but that judgments of curvedness were inaccurate and highly variable (see also Experiment 4 of Norman et al. (2006)). Numerous shape perception studies have found that metric 3D structure is not perceived accurately, including studies of structure from motion (e.g., Norman & Lappin, 1992; Norman & Todd, 1993; Perotti et al., 1998; Tittle et al., 1995; Todd & Bressan, 1990; Todd & Norman, 1991), or of binocular stereopsis (e.g., Johnston, 1991; Tittle et al., 1995), or of the combination of binocular disparity and motion (Tittle & Braunstein, 1993; Tittle et al., 1995), and even of structure under relatively full cue conditions (Norman & Todd, 1996; Norman, Todd, & Phillips, 1995).

These results are consistent with those in studies of object recognition where observers have been found not able to use metric

* Corresponding author. Address: Department of Psychology, Indiana University, 1101 East Tenth Street, Bloomington, IN 47405-7007, USA. Fax: +1 8128554691.
E-mail address: gbingham@indiana.edu (G.P. Bingham).

0042-6989/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.visres.2012.07.013

shape to recognize objects effectively. Instead, observers have been thought to use qualitative object properties for recognition. There have been two competing theories about object recognition. According to viewpoint-dependent theories, a set of image-based 2D templates of an object viewed from various perspectives generalizes to perception of the whole object (Christou & Bülthoff, 2000; Edelman & Bülthoff, 1992; Hayward & Tarr, 1997; Tarr, 1995; Tarr & Bülthoff, 1995; Tarr et al., 1998). Edelman and Bülthoff (1992) argued that templates are created for each experienced viewpoint and each view is stored in memory. When objects, such as a set of bent paper clips, were viewed at a different perspective in depth from that previously experienced (i.e., when the view of objects was not familiar), reaction times and error rates increased during recognition.

In contrast, other theorists have argued for viewpoint invariance in object recognition (Biederman, 1987, 2001; Biederman & Bar, 1999, 2000; Biederman & Gerhardstein, 1993, 1995). Viewpoint-invariance theory assumes that objects can be recognized from unfamiliar or inexperienced viewpoints. This should be possible, according to this theory, when objects differ in non-accidental properties (NAPs) of parts, i.e., 'geons'. Biederman and Gerhardstein (1993) described three conditions required to support viewpoint-invariance in recognizing objects. First, the object must be composed of easily identifiable parts whose shape is qualitatively distinct. Second, the object must have different geons in similar configurations (e.g., a curved-axis cylinder vs. straight cylinder on top of a brick) or similar geons in different configurations (e.g., a curved-axis cylinder on top of a brick vs. next to a brick) so as to make qualitatively distinct configurations. Third, visible parts of the object should not be suddenly changed over viewpoint changes, instead visible parts should be gradually occluded by rotated viewpoints.

According to viewpoint-invariance theory, objects are recognizable using their distinct qualitative properties. NAPs are distinguished from quantitative (metric) properties like the aspect ratio of a part or the degree of curvature of a contour. Although the shape perception literature seemed to show that observers cannot perceive metric shape well, it remained possible nevertheless that metric shape is used for object recognition. Biederman and Bar (1999) investigated how well objects are discriminated from different views in depth when they have different qualitative properties (NAPs) or when they have different quantitative (metric) properties (MPs) (see also Vogels et al. (2001)). They created 12 sets of computer generated two-part 3D novel objects. Each set included an original object, an object that is qualitatively different from the original one (that is, NAP or geon change), and an object quantitatively different from the original one (that is, MP change). For example, an object was constructed of two cylinders attached perpendicularly to one another at an elbow, and the MP change was a change in the size of the cylinders whereas the NAP change was a change from cylindrical to octagonal shape. Each object was depicted at two orientations in depth. Observers judged whether a sequential pair of objects was same or different, ignoring differences of orientation in depth. The results were different depending on whether the difference between the objects was qualitative (NAPs) or quantitative (MPs). When the viewpoint was rotated in depth, the ability to detect the quantitative differences (MPs) was poor (the correct judgment rate was at chance) whereas the ability to detect qualitative differences (NAPs) was good. Novel objects differing qualitatively (NAPs) were discriminated much faster and more accurately than objects differing quantitatively (MPs). This result was consistent with viewpoint-invariance theory in which NAP differences have a privileged role in object recognition.

More recently still, Foster and Gilson (2002) essentially replicated these results using a discrimination task. They reported that recognition based on metric properties was three times worse than performance based on the number of parts in an object, that is, when objects differed qualitatively.

As we noted, the finding that only qualitative (or NAP) differences are well recognized is consistent with shape perception studies finding that metric shape cannot be perceived accurately. An inability to perceive metric shape accurately would not only place limits on object recognition. It would also be a problem for actions like visually guided reaches-to-grasp. Grasps are known to be accurately sized relative to an object as the hand approaches the object. When the grasp involves contact of thumb and fingers on the front and back of the object, respectively, then the specification of the relevant extent of the object (in depth) requires combined information about object size and shape. (See Lee et al. (2008, 2010) for discussion and illustration of this point.) People do perform accurate reaches-to-grasp despite the apparent inability to perceive metric shape accurately. Lee et al. (2008) investigated whether haptic feedback would calibrate shape perception in the context of reaches-to-grasp allowing it to be metrically accurate. Lee et al. found that shape perception is not calibrated by haptic feedback information. Accordingly, they suggested that this is the reason that on-line guidance is frequently used in reaches-to-grasp.

However, contrary to previous results on perception of metric shape, Bingham and Lind (2008) showed that it is possible to perceive metric structure accurately given large perspective changes, namely a continuous 45° change or rotation. Observers judged the front, back and sides of virtual objects using a stylus to touch the locations. Objects were rotated or observers actively moved around objects. Observers perceived the metric shape of 3D objects correctly when the perspective was rotated either passively or actively by 90°. In addition, Bingham and Lind asked observers to judge the shape of objects after the objects were rotated by either 30°, 45°, 60°, or 90° to test whether as much as 90° rotation was necessary for accurate shape perception. They found that the judgment of the object shape was incorrect with only 30° rotation but that metric structure could be perceived with at least 45° rotation. Observers also attempted to judge shape from two discrete views differing by 90° to investigate whether this was enough to perceive metric shape. The judgments were not accurate. This is consistent with the results of Biederman and Bar (1999) in which observers were poor in detecting the quantitative (metric) differences from two discrete views.

According to the results of Bingham and Lind (2008), a continuous 45° perspective transformation is both necessary and sufficient to perceive metric shape accurately. Lee, Lind, and Bingham (2008) also found that metric 3D shape can be perceived accurately from continuous 45° perspective change. 3D cylindrical objects were shown, in a computer graphics stereo (anaglyph) display, rotating by different amounts. The amount of continuous rotation was parametrically varied from 11.5° up to 45°. Following each display, a 2D ellipse appeared on the computer screen. By pressing computer keys, observers adjusted the shape of the 2D ellipse so as to match the cross-section of the 3D cylindrical object. Accurate judgment of the metric shape was achieved only with 45° perspective change. This result confirmed the finding that accurate metric shape perception can be achieved using information from large perspective change, 45° or greater, but not less.

Now the question is: If such information from large perspective change is available, is it possible to recognize objects using metric properties? The reason that observers were unable to recognize objects using metric properties in previous object recognition studies might not be because we can detect only qualitative properties (or NAPs), but because sufficient information simply was not provided. Thus, in this study, we provided information from either small or large continuous perspective change, and investigated whether such information yielded the ability to detect quantitative properties and use them to recognize objects. Relatedly, we also now investigated whether the previous results, found using cylindrical objects, generalize to other shapes, namely polygonal and asymmetric shapes relevant to general object recognition abilities. Previously, we found that metric shape could be perceived and used to guide reach-to-grasp actions. In the context of the Two Visual System theory of Goodale and Milner, this result confirmed the ability to detect metric object properties as expected of the dorsal or perception-for-action system (Hu & Goodale, 2000; Milner & Goodale, 1995, 2006, 2008). The central function of the ventral system is object recognition. In the context of this theory, we now test whether metric shape can be detected and used also by the ventral system. As described in the General Discussion, the theory predicts that the ventral system should perceive metric shape, but this has not been previously confirmed.

2. Methods

2.1. Participants

Twenty adults aged 18–50 participated as observers in this experiment. Ten observers, three males and seven females, participated in a small rotation condition and ten, four males and six females, participated in a large rotation condition. All had normal or corrected to normal vision and passed a stereo fly test (Stereo Optical Co., Inc.) that was used to check stereoscopic depth perception. All of the participants were naïve as to the purpose of the study and were paid at $7 per hour. All procedures were approved by and conform to the standards of the Indiana University Human Subjects Committee.

2.2. Stimuli and apparatus

Twenty-four octagonal objects were used with cross sections as illustrated in Fig. 1. There were four qualitatively different objects and each object had five quantitatively different variations. Four qualitatively different objects were constructed by varying the number of concave vs. convex vertices from 0 and 8, respectively, to 3 and 5, respectively. Each quantitatively different object was compressed or stretched in -10% or $+10\%$ steps from a symmetrical octagon, yielding aspect ratios of 0.8, 0.9, 1.0, 1.1, 1.2, and 1.3 (and labeled 1–6 in Fig. 1). Three-dimensional (3D) objects were generated in a computer display using red–blue anaglyphs to create stereo. The red and blue were calibrated in a pilot study to eliminate cross talk. A representative object is illustrated in Fig. 2 using a standard stereogram with crossed disparity. The two-dimensional (2D) objects shown in Fig. 1 appeared as white octagonal shapes on a black background, each with a cross-sectional shape corresponding to one of the 3D objects. Each stimulus was displayed on a Mitsubishi Diamond Plus 74SB CRT computer screen with a resolution of 1280×1024 and a frame rate of 60 Hz. Using a chin rest, the cyclopean eye of the observer was fixed at 55 cm from the screen and 12 cm above the bottom of the object on the screen.

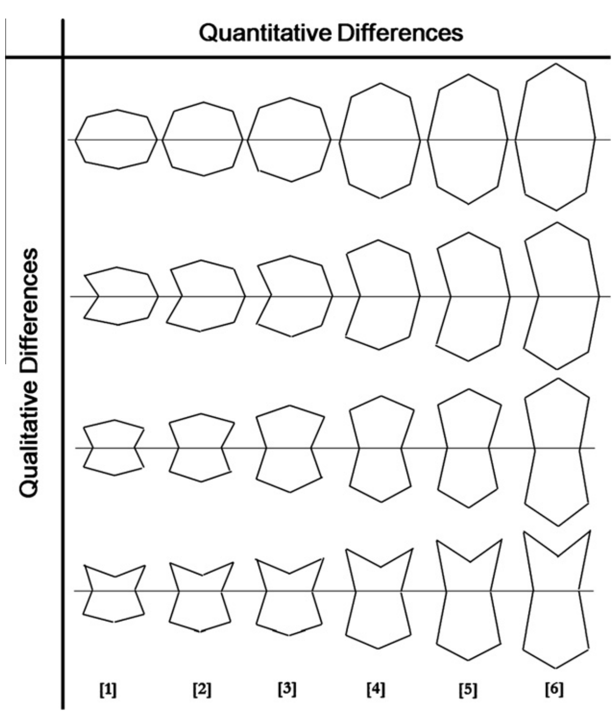

Fig. 1. A schematic representation of the cross-sectional shape of the objects used in this study.

2.3. Experimental design and task

The study was designed as a four factor mixed design with task (2D quantitative, 3D qualitative or 3D quantitative), amount of difference and repetition as within-subjects factors and rotation amount (small or large) as a between-subjects factor. The observers were randomly assigned to small (10Ss) and large (10Ss) rotation groups. There were three tasks for each observer: a 2D quantitative difference task and two 3D tasks, (1) 3D quantitative difference and (2) 3D qualitative difference. We tested the 2D task for two reasons. First, we needed to tune observers to the 3D quantitative difference task to be sure that observers understood which dimension of difference they were to judge. Second, we needed to determine the level of performance in the easiest 2D case to provide some basis for evaluation of the performance in the 3D task.

2.3.1. 2D quantitative difference task

Every possible pair of quantitatively different objects was used in an XAB paradigm for the 2D task. The task was started by pressing a key causing a 2D target object to be shown. After the 2D target object, the following question appeared: ‘which of the following two objects is the same as the target object?’. Then, a pair of the 2D objects was shown successively with a 1.6 s delay (black screen) between the two objects. One of the pair of objects was the same as the test stimulus, and the other was a probe stimulus. The probe stimulus was one of 5 quantitatively different (but qualitatively similar) objects or an object that was the same as the target stimulus for a catch trial. The order of test and probe stimuli was randomized. After the pair of objects was shown, the instruction; ‘press ‘a’ if the first one is the same and press ‘l’ if the second one is the same’ was shown. The observer had to judge as accurately and quickly as possible right after the instruction was shown. Observers pressed the space bar to begin each trial so trials were self-paced. Each observer judged a total of 144 trials (4×36, that is, 36 comparisons for each of the 4 qualitatively distinct types, where in each of these four types, each of the 6 quantitatively different objects was compared to itself and the 5 others yielding $6 \times 6 = 36$ comparisons) in a random order.

2.3.2. 3D tasks

The 3D tasks included a 3D quantitative difference task and a 3D qualitative difference task. The 3D target object was shown oscillating back and forth. It disappeared and then a 2D test object was shown. The observer had to judge whether the 2D test object was the same in shape as the cross-section of the 3D target object. Judgments were to be as accurate and quick as possible. Two rotation amounts were tested as a between-subject factor yielding

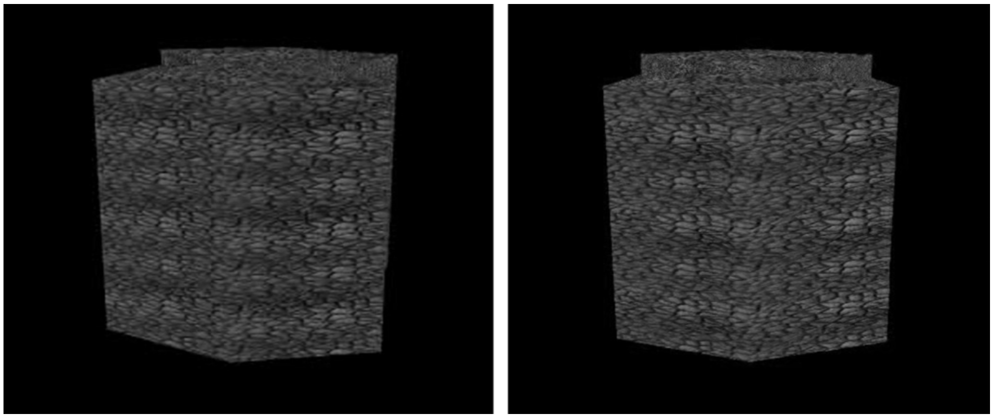

Fig. 2. A stereogram (crossed disparity) of a sample 3D object used in these studies. Actual displays consisted of red and blue textures.

large and small rotation conditions. In the large rotation condition, the 3D target object was rotated by 35° first to one side from a canonical view (that is, looking down the longer principle axis) and then to the other side so that the perspective was changed by a total of 70°. The object was rotated in this way twice and it took around 5 s to finish rotating. In the small rotation condition, the 3D target object was rotated by 10° first to one side from a canonical view and then to the other side so that the perspective was changed by a total of 20°. The object was rotated in this way three times and it took around 4.6 s to finish rotating. Each 3D task included a practice session and an experimental session. The practice session proceeded the experimental session and only the large rotation amount was tested in the practice session. In the practice session, there were 12 trials in a block. The observers performed the judgment and then feedback was provided. The block of 12 practice trials was repeated until the observers judged correctly more than 10 times within a block or they judged 30 trials in total. Then, they proceeded to the experimental session. No feedback was provided in the experimental session.

2.3.3. 3D quantitative difference task

The 3D target object and the 2D test object in a trial differed quantitatively, that is, in aspect ratio. In the practice session, we used objects with 4 concave and 4 convex edges that were not used in the experimental session. The 3D target object was always the most compressed (0.8 in the *z*-direction) and the 2D test object was either the same or one of 5 quantitatively different objects (0.9, 1.0, 1.1, 1.2 or 1.3 in the *z*-direction). Thus, a total of 12 trials in a practice block consisted of 6 trials (i.e., 1 same and 5 quantitative difference trials) and 2 repetitions.

In the experimental session, only either the most compressed (0.8 in the *z*-direction, i.e., object [1]) or the most expanded objects (1.3 in the *z*-direction, i.e., object [6]) in the range of quantitative variations were used for the 3D target object. Each 3D target object had five difference trials and 1 same trial. The observer performed a total of 144 trials in a random order (8 objects × 6 same/different trials × 3 repetitions).

2.3.4. 3D qualitative difference task

The 3D target object and the 2D test object in each trial differed qualitatively. In the practice session, we used 3D target objects that were qualitatively different from the objects used in the experimental session, namely, objects with 4 concave vertices (not shown in Fig. 1). A total of 12 trials in a practice block consisted of 4 trials (1 same and 3 qualitative difference trials using the qualitatively different objects with the same aspect ratio, that is, from the same column in Fig. 1) and 3 repetitions. In the experimental session, all 24 objects were used for the 3D target object and each object had 1 same and 3 qualitative difference trials. Each observer performed a total of 192 trials in a random order (24 objects × 4 same/different trials × 2 repetitions).

2.4. Procedure

The participants read and signed the consent forms, and then they underwent the stereo fly test to check stereoscopic depth perception. The test includes a graded circle test (ranging from 800 to 40 s of arc). Four circles appear within each of 9 squares and only one of the circles has a degree of crossed disparity. The participants were asked to indicate which circle appears raised. They had to answer correctly for at least 6 of the 9 levels to pass this test. This represents a stereo acuity of 80° per arc sec. If they passed the stereo test, then they adjusted the height of a chair in the testing room to fit a chin rest comfortably.

Each participant performed three test sessions in the same order, namely, the 2D quantitative difference, 3D quantitative difference and 3D qualitative difference tasks. The experimenter explained that objects would differ quantitatively in the first two sessions and qualitatively in the last session. Before each test session, the task and the procedure were described to the participant. Each trial was begun by pressing a spacebar. This allowed the observers to take a break anytime they needed during the session. Participants hit one of two keys on a keyboard to judge whether the 2D object was the same or different from the preceding 3D object. They were instructed to do this as quickly and accurately as possible. The judgment and reaction time data were saved for the analysis. The observers were debriefed and compensated after the sessions were completed.

3. Results

We analyzed the data of each task separately. The design of this study was complicated. Fig. 3 illustrates this design and the comparisons that we performed and the results. d' were computed and used together with proportion correct to evaluate accuracy. Both d' and RT were compared across task conditions as shown. Effects or factors that failed to reach significance were $p > 0.1$ or greater in all cases.

3.1. Analysis of judgment accuracy

3.1.1. 2D quantitative difference task

All 2D quantitative differences were readily recognized, but the smallest differences were not as well recognized as the larger ones.

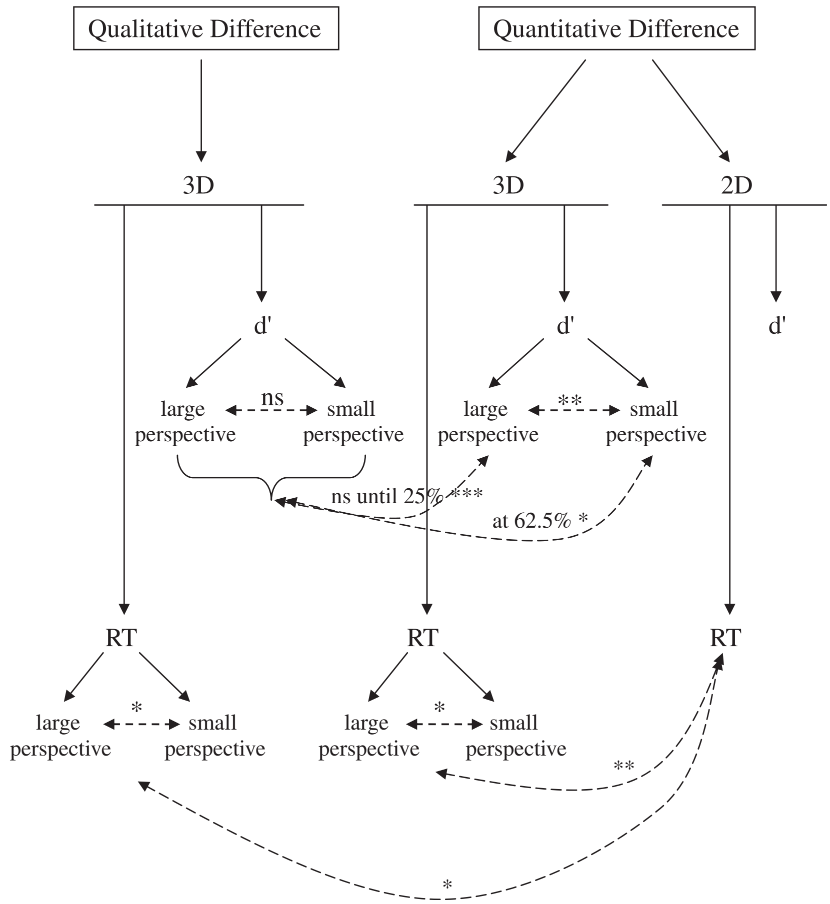

Fig. 3. A diagram illustrating the comparisons made among the different conditions and experiments for the two types of measures, d' and RT. The dashed lined represent comparison performed using ANOVA: $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$.

Fig. 4 shows both the proportion of correct judgments and the d' plotted as a function of the amount of difference between the target and the probe object. The proximity of the target and probe objects within the range of quantitative variations was used to determine the quantitative difference. Thus, the amount of difference varied from 0 to 5. If the target and the probe object were neighboring (e.g., object [3] and object [4]), the amount of difference was 1 and indicated that the quantitative difference between two objects was small. Participant's judgments were not biased as shown by the proportion of correct judgments (≈0.5) for 'same' trials in which the two objects following the target object were the same. A mixed design ANOVA on correct judgments with quantitative difference (adjacency of objects) as a within subject factor and group (large vs. small rotation condition) as a between subject factor yielded a significant effect of quantitative difference, $F(4,72) = 40.5$, $p < .001$, $\xi^2 = 0.67$. The ANOVA did not yield either a main effect of group or a significant interaction. The 2D quantitative difference task was same in the large and small rotation conditions, and observers performed equivalently in both conditions. (There was no rotation in the 2D task. We tested the two groups just to be sure they were otherwise homogeneous.) As shown in Fig. 4a, participant's judgments reached about 80–90% correct once differences were 2 or greater.

We computed d' with respect to quantitative difference (see Fig. 4b). We could not compute d' for the largest quantitative difference (i.e., when the adjacency level was 5) and thus excluded this condition from analysis. A mixed design ANOVA on d' with quantitative difference (adjacency of objects) as a within subject factor and group (large vs. small rotation condition) as a between subject factor yielded a significant effect of quantitative difference [$F(3,54) = 45.7$, $p < .001$, $\xi^2 = 0.70$], but neither a main effect of group nor a significant interaction. Even though d' was lower for an adjacency level of 1, the d' was already equal to 1 or greater indicating relatively good performance. Nevertheless, the results showed that detecting quantitative changes even among 2D objects is more difficult if they differ quantitatively by a small amount.

3.1.2. 3D quantitative difference task

The results showed that large perspective changes allowed objects with modest differences in aspect ratio (≈20–25% or greater) to be recognized reliably while small perspective changes only allowed objects with large differences (≈70%) in aspect ratio to be so recognized. We plotted mean d' with respect to the quantitative difference between a pair of objects as shown in Fig. 5. We used only object [1] and object [6] as the 3D target objects. The adjacency level was computed by subtracting the 2D test object number from the 3D target object number. Thus, object [1] had negative adjacency levels and object [6] had positive adjacency levels indicating the quantitative difference between two objects. When the quantitative difference between a pair of objects was large (e.g., adjacency level was 4 or 5), sometimes hit rate was 1.0 and miss rate was 0 and thus, we were unable to calculate a z value to compute d'. In this case, we replaced 1 with 0.95 (1–1/2N) for hit rate and 0 with 0.05 (1/2N) for false alarm rate to calculate d'. As can be seen in Fig. 5a, observers discriminated objects better when the 3D target object was presented with large perspective changes than with small perspective changes. However, the difference between

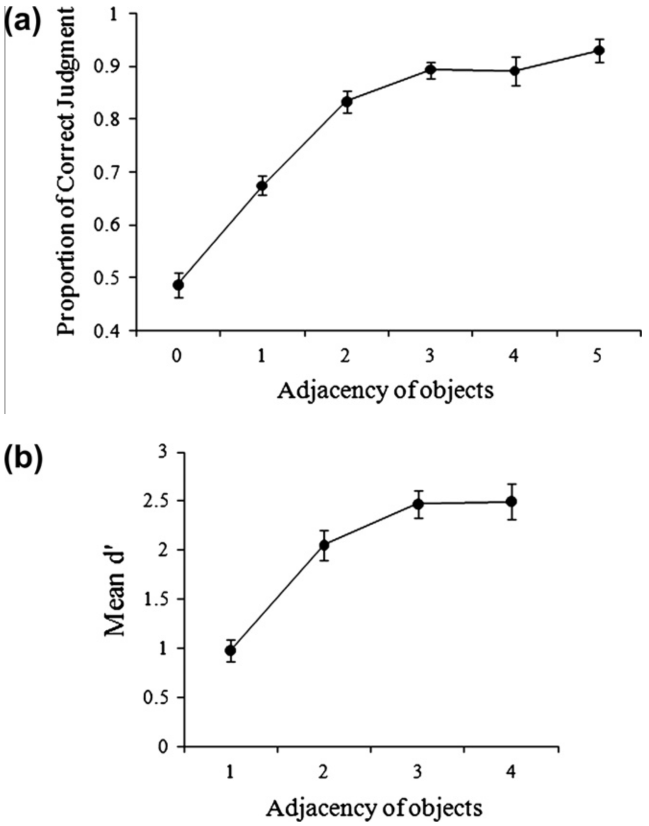

Fig. 4. Results of the 2D task: (a) mean percentage and (b) mean d' plotted with respect to adjacency of objects, namely the absolute value of the difference of the two object numbers. If the number is larger, the two compared objects were more different.

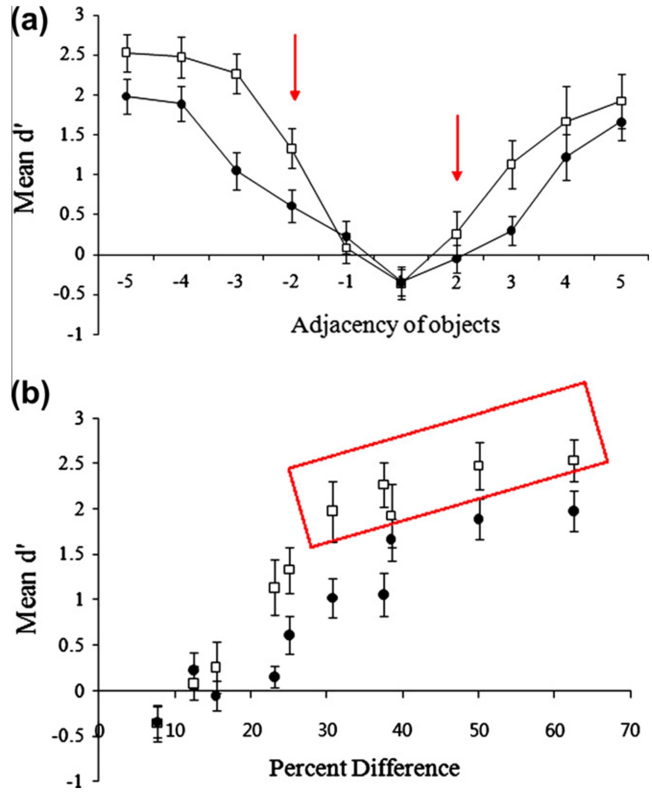

Fig. 5. Mean d' of the 3D quantitative task plotted with respect to (a) adjacency of compared objects and (b) absolute percent difference in aspect ratio of the compared objects. The open squares represent the large rotation condition and the filled circles represent the small rotation condition.

the results for large perspective changes and those for small perspective changes was larger when the 3D target object was object [1] than when it was object [6] at the same adjacency level. For instance, when the adjacency level was |2| in the large rotation condition, mean d' for object [1] was 1.33 whereas for object [6] it was 0.26. (See the downward pointing arrows in Fig. 5a.) The proportions of quantitative difference were different depending on comparison object size in depth. Therefore, we computed the percent differences in terms of object depth calculating percent of quantitative difference with respect to the 3D target object. The percentages of quantitative difference were 12.5%, 25%, 37.5%, 50%, and 62.5% for the compressed object (object [1]) and 7.7%, 15.38%, 23.08%, 30.77%, and 38.46% for elongated object (object [6]), respectively. We plotted mean d' with respect to the percentage of quantitative difference (see Fig. 5b) with the result that mean d' ordered as a function of the percentages. We performed a mixed design ANOVA on d' with percentage of quantitative difference as a within subject factor and rotation amount (large vs. small) as a between subject factor. We found a significant effect of percentage of quantitative difference [$F(9,162) = 57.1$, $p < .001$, $\xi^2 = 0.73$], a significant effect of rotation amount [$F(1,18) = 4.63$, $p < .05$, $\xi^2 = 0.20$], and a significant interaction [$F(9,162) = 2.8$, $p < .005$, $\xi^2 = 0.04$]. Optical information from large perspective changes allowed 3D objects to be recognized using modest differences in quantitative properties and optical information from small perspective changes did not.

3.1.3. 3D qualitative difference task

The results showed that observers improved in their ability to discriminate qualitatively different 3D objects as the number of

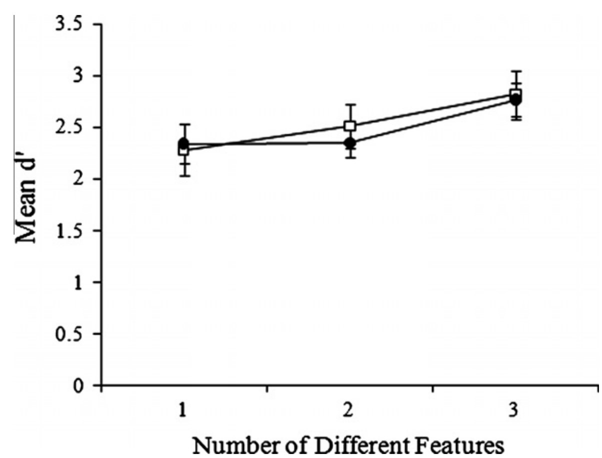

Fig. 6. Mean d' of the 3D qualitative task plotted with respect to the number of different features. The open squares represent the large rotation condition and the filled circles represent the small rotation condition.

qualitatively different features (i.e., number of convex vs. concave vertices) increased but they were able to recognize objects well even with a small amount of difference and with small perspective changes. As shown in Fig. 6, observers recognized 3D objects correctly across all qualitative differences. A mixed design ANOVA on d' with qualitative difference (number of different features) as a within subject factor and rotation amount (large vs. small) as a between subject factor yielded a significant main effect of qualitative difference [$F(2,36) = 10.47$, $p < .001$, $\xi^2 = 0.34$] but neither a main effect of rotation amount nor a significant interaction.

3.1.4. 3D quantitative difference task vs. 3D qualitative difference task

We investigated whether information from large perspective changes rendered recognition performance based on quantitative shape differences comparable to performance based on qualitative differences (see Figs. 5b and 6). The results showed that, for small perspective changes, the ability to detect quantitative differences was significantly and consistently poorer than the ability to detect qualitative differences. For large perspective changes, however, the ability to detect quantitative differences was equivalent to the ability to detect qualitative differences once the percentage of quantitative shape difference exceeded 25%.

We compared mean d' at each percentage of quantitative difference for the large rotation and small rotation conditions separately with mean d' of all qualitative conditions combining the two rotation conditions (because they yielded no differences). We performed two-tailed independent two sample *t*-tests (for which the two sample sizes are unequal and the variance is assumed to be different for each case). In the small rotation condition, the mean d' of even the largest percentage of quantitative difference, 62.5%, was significantly lower than the mean d' of the qualitative task [$t(12.194) = 2.3$, $p < .04$]. However, in the large rotation condition, as percentage differences decreased, the quantitative task was different from the qualitative task only at a percentage difference of 25% or less. The mean d' for 25% of quantitative difference was significantly lower than the mean d' of the qualitative task [$t(11.442) = 4.5$, $p < .001$].

3.2. Analysis of RT

3.2.1. 2D quantitative difference task

Although judgments varied with respect to the quantitative difference, analyses of reaction times did not show any significant effects. The mean RT was 1.14s.

3.2.2. 3D quantitative difference task

The analysis of d' had showed two things. First, with small perspective changes, performance was always inferior to that with large perspective changes as well as to that making qualitatively based judgments. Second, with large perspective changes, once the percentage quantitative difference was larger than 25%, performance was equivalent to that making qualitatively based judgments. We expected reaction time results to reflect these differences and in particular, we expected a significant change in RT once the percent quantitative difference was greater than 25%, at which point RTs for recognition of quantitative differences should be the same as RTs for recognition of qualitative differences. This is exactly what happened.

As shown in the left panel of Fig. 7, there was discrete change of mean reaction times in the large rotation condition between a 25% and 30.77% quantitative difference and performance was flat on either side of this jump. (See the boxed set of means for larger differences that compare with the mean of the means for qualitative differences as shown by the dotted line.) No such change occurred with small perspective changes. Given this, we performed a mixed design ANOVA on reaction times with percentage of quantitative difference as a within subject factor and rotation amount (large vs. small) as a between subject factor, separately for the first portion of percentage differences (7.7–25%) and then, for the second portion of percentage differences (30.77–62.5%). While the ANOVA for the first half of percentage differences (7.7–25%) did not yield any significant effects, the ANOVA for the second half of percentage differences (30.77–62.5%) yielded a significant effect of percentage of quantitative difference [$F(4,72) = 3.4$, $p < .02$, $\xi^2 = 0.29$], a significant effect of rotation amount [$F(1,18) = 7.0$, $p < .02$, $\xi^2 = 0.15$], and no interaction.

3.2.3. 3D qualitative difference task

The results showed that observers detected qualitative differences faster as the objects differed by more features and faster with large perspective changes. As shown in the right panel of Fig. 7, reaction times revealed a difference between the large and the small rotation condition. A mixed design ANOVA on reaction times with qualitative difference (number of different features) as a within subject factor and rotation amount (large vs. small) as a between subject factor yielded a main effect of qualitative difference [$F(2,36) = 12.6$, $p < .001$, $\xi^2 = 0.38$], and a main effect of rotation amount [$F(1,18) = 5.2$, $p < .04$, $\xi^2 = 0.22$], but no interaction.

3.2.4. 2D quantitative vs. 3D quantitative vs. 3D qualitative task

As can be seen in Fig. 7, observers were able to recognize 2D objects faster than 3D objects no matter what properties were to be compared (quantitative or qualitative) in 3D objects. We performed two two-factor ANOVAs on reaction times with group (2D quantitative vs. 3D quantitative or 3D qualitative, respectively) and rotation amount (large vs. small) as factors. The ANOVAs yielded a significant effect of group for 2D quantitative vs. 3D quantitative [$F(1,18) = 45.0$, $p < .001$, $\xi^2 = 0.20$] and for 2D quantitative vs. 3D qualitative [$F(1,18) = 4.8$, $p < .05$, $\xi^2 = 0.70$] but neither an effect of rotation nor an interaction. As compared to 2D recog-

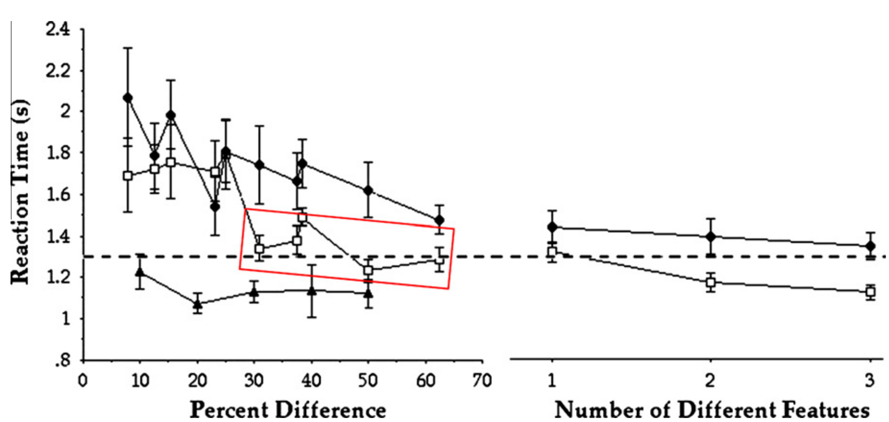

Fig. 7. Reaction time results for all tasks and experiments. The left panel shows reaction times for the 2D task and the 3D quantitative task and the right panel shows reaction time for the 3D qualitative task. The open squares represent the large rotation condition and the filled circles represent the small rotation condition. The filled triangles represent the 2D task. In this case, 10–50 represents adjacency of objects, 1–5. The dashed line represents the mean reaction time of the 3D qualitative task combining the large and the small rotation condition.

nition tasks, performing recognition of 3D objects, whether based on qualitative or quantitative properties, incurred a cost in reaction time.

As mentioned earlier, there was a discrete change in performance of the 3D quantitative task given large perspective changes once the percentage of quantitative difference exceeded 25%. At this point, the reaction times were comparable to the mean of the reaction time for detecting qualitative differences (represented by the dashed line in Fig. 7). An ANOVA comparing the subject RT means for quantitative large perspective change and differences greater than 25% vs. qualitative changes was not significant ($p > 0.7$). The same ANOVA comparing quantitative small perspective changes to qualitative changes was significant, $F(1,9) = 15.6$, $p < 0.005$, $\xi^2 = 0.64$. This finding was consistent with the finding that once the percentage of quantitative shape difference exceeded 25%, recognition performance based on quantitative shape differences with large perspective changes was comparable to performance based on qualitative differences. Thus, once information from large perspective changes is available, metric shape perception is possible and so is the ability to recognize objects that differ only quantitatively. Further, performance in recognizing quantitatively distinct objects is comparable to performance in recognizing qualitatively distinct objects.

4. Discussion

Human observers typically are able to recognize familiar objects even when they are viewed at arbitrary orientations in depth. The question is how are we able to do this. View-dependent theory has suggested that we experience image-based 2D templates of the object from various orientations and store them in memory to generalize to the whole object (Christou & Bülthoff, 2000; Edelman & Bülthoff, 1992; Hayward & Tarr, 1997; Tarr, 1995; Tarr & Bülthoff, 1995; Tarr et al., 1998). In contrast, viewpoint invariance theory has suggested that we are able to recognize objects from a novel orientation in depth without much experience of different orientations. This letter theory, however, assumes that novel objects can be recognized from previously un-experienced viewpoints only when non-accidental properties (NAPs) of parts (i.e., geons) can be exploited (Biederman, 1987, 2001; Biederman & Bar, 1999, 2000; Biederman & Gerhardstein, 1993, 1995). According to the theory, novel objects can be recognized under rotation when objects differ in qualitative properties (NAPs), but not when they differ only in quantitative (metric) properties (MPs). Biederman and Bar (1999) found that novel objects differing qualitatively (NAPs) were discriminated much faster and more accurately than objects differing quantitatively (MPs) when they were seen at different orientations in depth. Foster and Gilson (2002) also found that different objects were discriminated much more accurately when they differed qualitatively rather than metrically. This finding of a relative inability to detect the metric differences is consistent with numerous previous perception studies in which observers have been unable to perceive metric 3D shape accurately (Johnston, 1991; Norman & Lappin, 1992; Norman & Todd, 1993, 1996; Norman, Todd, & Phillips, 1995; Perotti et al., 1998; Tittle & Braunstein, 1993; Tittle et al., 1995; Todd & Bressan, 1990; Todd & Norman, 1991).

However, the inability to perceive metric shape accurately would be a problem in the context of visually guided actions like reaches-to-grasp. Control of grasping typically requires combined information about object size and shape because the grasp involves contact of thumb and fingers on the front and back of the object, respectively. We typically do not have any problem performing accurate reaches-to-grasp despite this apparent inability to perceive metric shape accurately. Lee et al. (2008) found that poor shape perception yielded inaccurate feedforward grasping and accordingly, inferred that this is the reason on-line guidance is often required for accurate grasping.

However, Bingham and Lind (2008) found that metric structure can be perceived accurately from large perspective changes and used to guide accurate reaches. Lee et al. (2008) then found that this information could be used to guide accurate grasping as well. Both reaching and grasping are dorsal system tasks. The dorsal system is hypothesized to be able to detect and use metric object properties and indeed, the findings of accurate perception of both metric shape and metric size used to guide reaching and grasping, respectively, support this expectation. We now investigated whether the ventral system, hypothesized to be responsible for object recognition abilities, might be able to detect and use the same information to perceive metric shape and use it to recognize objects. According to the Two Visual System theory, the dorsal system is able to perceive absolute metric object properties (i.e. Euclidean structure) whereas the ventral only perceives relative object properties. In the Klein hierarchy of geometries, this would entail properties captured by Similarity geometry. Similarity geometry allows a single isotropic free scale factor not allowed by Euclidean geometry. This means that the ventral system would not perceive metric size, but it should perceive metric shape and use it for the primary function of the ventral system, namely, object recognition. Although predicted by the Two Visual System theory, this ability has not previously been demonstrated. In fact, previous results from shape perception and object recognition studies showed that properties captured only by Affine geometry were accurately perceptible. Affine geometry introduces a second free scale factor making it impossible to perceive relations between metric scale in different directions, and thus, impossible to perceive metric shape. (See Lee and Bingham (2010) and reference therein cited for additional discussion and explanation.)

When information from large perspective changes is made available, previous studies involving action measures showed that accurate metric shape perception results. This suggested that we should return to the object recognition problem to investigate whether the reason observers were unable to recognize metric properties in previous studies was simply because the available information was not sufficient for detecting metric shape. Thus, we now investigated whether 3D objects varying only in metric shape could be recognized by the ventral system when information from large continuous perspective changes was available. In addition, we tested polygonal and asymmetric shapes that would be relevant to more general object recognition tasks.

We found that given large perspective changes, observers were able to detect quantitative properties of objects to perform fast and accurate object recognition. This result, for the first time, confirmed the original hypothesis of the Two Visual System theory that the ventral system should be able to detect and use metric shape properties captured by Similarity geometry.

In the 2D quantitative difference task, observer's judgments of quantitative properties of 2D objects were generally good as measured by d' and the amount of quantitative differences. However, we noted that it remained more difficult to detect quantitative shape differences even in 2D objects if quantitative differences were small. Reaction times, however, were faster in the 2D task than in any of the 3D tasks, even the 3D qualitative difference task. So, 3D recognition tasks entail a cost in reaction time as compared to otherwise comparable 2D recognition tasks.

In the 3D quantitative difference task, we found that large perspective changes yielded the ability to detect quantitative properties, while small perspective changes did not. Observers were able to recognize quantitative differences with large perspective changes once the percentage difference of quantitative properties exceeded 25%. At this point, observers performed accurately and quickly in detecting quantitative properties. In the 3D qualitative difference task, observers' recognition judgments were good in general for all differences of qualitative properties. Observers, however, discriminated objects better and faster as the difference in qualitative properties was larger (meaning objects had more qualitatively distinct features). Furthermore, we found that there was a significant effect of rotation amount in which observers discriminated qualitatively different objects faster with large perspective changes than with small perspective changes. Thus, although large perspective changes were not necessary to recognize objects using qualitative properties, if information from large perspective changes was available, observers exploited this information to recognize objects more rapidly.

The more important question is whether information from large perspective changes rendered performance in recognizing quantitative differences comparable to performance in recognizing qualitative differences. Observers were able to discriminate quantitative differences with large perspective changes once objects differed in quantitative properties by more than 25% and that ability was equivalent to the ability to detect qualitative differences. In contrast, the ability to detect quantitative differences with small perspective changes was poor, even when the percentage of quantitative differences was large, 62.5%. Reaction time results were consistent with these findings for judgment accuracy. Reaction times for detection of quantitative properties with large perspective changes were significantly less than those with small perspective changes once the percentage of quantitative difference exceeded 25%. At this point, the reaction times to detect quantitative properties were comparable to the reaction times to detect qualitative properties.

The conclusion is that metric properties of objects can be detected and used for recognition as predicted of the ventral system by the Two Visual System theory. A last consideration might be to wonder when the large perspective changes (≈45°) required for perception of metric shape might normally be available to an observer. Such changes are commonly available when we locomote through the environment. When one enters one's kitchen in the morning or one's office at the beginning of the workday, the objects arrayed in the room are typically viewed with such large perspective changes. Lee and Bingham (2010) investigated whether this information would be effective later in time to support successive reaches-to-grasp the different objects initially viewed as arrayed about such an environment. The results demonstrated that the information provided by the initially available larger optic flows persists in the subsequently available image structure.

References

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*, 66–80.

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

Biederman, I. (2001). Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vision, 13*, 241–253.

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

Biederman, I., & Bar, M. (2000). Differing views on views: Response to Hatward and Tarr (2000). *Vision Research, 40*, 3901–3905.

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.

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.

Bingham, G. P., & Lind, M. (2008). Large continuous perspective transformations are necessary and sufficient for accurate perception of metric shape. *Perception & Psychophysics, 70*, 524–540.

Christou, C., & Bülthoff, H. H. (2000). Perception, representation and recognition: A holistic view of recognition. *Spatial Vision, 13*, 265–275.

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

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society London B, 269*, 1939–1947.

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.

Hu, Y., & Goodale, M. A. (2000). Grasping after a delay shifts size-scaling from absolute to relative metrics. *Journal of Cognitive Neuroscience, 12*, 856–868.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research, 31*, 1351–1360.

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

Lee, Y., & Bingham, G. P. (2010). Large perspective changes (>45°) yield perception of metric shape that allows accurate feedforward reaches-to-grasp and it persists after the optic flow has stopped! *Experimental Brain Research, 204*, 559–573.

Lee, Y., Crabtree, C. E., Norman, J. F., & Bingham, G. P. (2008). Poor shape perception is the reason reaches-to-grasp are visually guided online. *Perception & Psychophysics, 70*, 1032–1046.

Lee, Y., Lind, M., & Bingham, G. P. (2008). Metric shape perception requires a 45° continuous perspective change. *Journal of Vision, 8*(6), 759a.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Milner, A. D., & Goodale, M. A. (2006). *The visual brain in action* (2nd ed.). Oxford: Oxford University Press.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia, 46*, 774–785.

Norman, J. F., & Lappin, J. S. (1992). The detection of surface curvatures defined by optical motion. *Perception & Psychophysics, 51*, 386–396.

Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics, 53*, 279–291.

Norman, J. F., & Todd, J. T. (1996). The discriminability of local surface structure. *Perception, 25*, 381–398.

Norman, J. F., Todd, J. T., Norman, H. F., Clayton, A. M., & McBride, T. R. (2006). Visual discrimination of local surface structure: Slant, tilt, and curvedness. *Vision Research, 46*, 1057–1069.

Norman, J. F., Todd, J. T., & Phillips, F. (1995). The visual perception of surface orientation from multiple sources of optical information. *Perception & Psychophysics, 57*, 629–636.

Perotti, V. J., Todd, J. T., Lappin, J. S., & Phillips, F. (1998). The perception of surface curvature from optical motion. *Perception & Psychophysics, 60*, 377–388.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2*, 55–82.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology, 21*, 1494–1505.

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

Tittle, J. S., & Braunstein, M. L. (1993). Recovery of 3-D shape from binocular disparity and structure from motion. *Perception & Psychophysics, 54*, 157–169.

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 663–678.

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 48*, 419–430.

Todd, J. T., & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics, 50*, 509–523.

Vogels, R., Biederman, I., Bar, M., & Lorincz, A. (2001). Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. *Journal of Cognitive Neuroscience, 13*, 444–453.