The contribution of qualitative and quantitative shape features to object recognition across changes of view

JEFFREY C. LITER
University of California, Irvine, California

Two experiments investigated the influence of qualitative and quantitative shape features on recognition of novel, four-component objects. Quantitatively different objects had different connection angles between the components. Qualitatively different objects had different connection angles and differently shaped components in some of the four positions. Old–new recognition declined less with changes of view for qualitatively different objects (Experiment 1). However, recognition of these objects was made to decline sharply with changes of view if subjects were biased to attend to the connection angles rather than the component shapes (Experiment 2), suggesting that the influence of different features depends on visual experience with those features. These results favor a feature-based model of shape representation that utilizes multiple feature types and that can rely on different features depending on particulars of the objects and the task.

The projected shape of an object can change dramatically if it is seen from different viewpoints. Nevertheless, with the exception of certain accidental viewpoints, object recognition typically seems effortless regardless of the viewpoint from which objects are seen. Theorists have proposed very different long-term representations of object shape in attempts to explain this phenomenon. Some theorists have argued that object shape is represented in long-term visual memory by part-based structural descriptions that are largely independent of the vantage point of the viewer (Biederman, 1987; Marr & Nishihara, 1978). According to these structural-description theories, viewpoint-invariant recognition is possible because different views of an object have the same structural description. Other theorists have argued that shape is represented in long-term memory by descriptions that are specific to the vantage point of the viewer (Edelman & Bülthoff, 1992;

Tarr, 1995; Tarr & Bülthoff, 1995; Tarr & Pinker, 1989). To achieve viewpoint-invariant recognition with viewpoint-dependent descriptions, either many descriptions of each object must be stored in long-term memory, each specific to a different viewpoint, or a transformation mechanism must be used to match different descriptions of the same object, or both.

These theories make different predictions about the recognition of novel views of previously seen objects. Theories that posit recognition via viewpoint-dependent descriptions predict poorer recognition of novel views than of previously seen views. According to these theories, descriptions of novel views do not match any previously stored descriptions, so novel views can be recognized only after a time-consuming or error-prone transformation mechanism has been applied. Theories that posit recognition via viewpoint-invariant structural descriptions predict that recognition will be equally good for novel and known views as long as the same structural description can be generated for both views.

The present study attempted to account for some of the discrepancies that have been reported in the object recognition literature by examining whether the types of features that are available to discriminate among a set of objects influences how easily the objects are recognized following a rotation in depth. It will be argued that researchers who have observed viewpoint-dependent recognition performance have used stimuli that must be discriminated by features that are not easily identified following a rotation in depth, whereas researchers who have observed viewpoint-independent performance have used stimuli that can be discriminated by features that are more easily identified in rotated views. Before describing the objects used in this study, it will be useful to review some of the previous research on viewpoint generalization.

Early investigations of viewpoint generalization with unfamiliar objects such as smoothly bent wires (Rock & DiVita, 1987; Rock, DiVita, & Barbeito, 1981), collections of cubes connected face to face (Tarr, 1995), or "paper clip" objects composed of cylinders connected end to end (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992) are consistent with viewpoint-dependent representations of object shape. Evidence collected by Tarr (1995; see also, Tarr & Pinker, 1989) and by Edelman and Bülthoff (1992) suggests further that the visual system stores multiple viewpoint-dependent descriptions of objects in long-term visual memory and that novel views of known objects are recognized by transforming stored descriptions (e.g., via mental rotation, Shepard & Cooper, 1982) or by interpolating among stored descriptions (Poggio & Edelman, 1990).

Biederman (1987; see also Hummel & Biederman, 1992) has argued that object shapes are represented in long-term visual memory by structural descriptions that specify the components making up the objects and the spatial relationships among the components. An important property of the components proposed by Biederman (i.e., "geons") is that their 3-D shapes can be inferred from their projected shapes from nearly any viewpoint. In principal, such a representation should allow for greater view generalization because the same structural description can be used to recognize an object from a wide range of viewpoints. Biederman and Gerhardstein (1993) proposed three conditions for viewpoint-invariant recognition: (1) It must be possible to describe the objects with geon structural descriptions (GSDs), (2) different objects must have different GSDs, and (3) different views of the same object must have the same GSD. According to Biederman and Gerhardstein, the researchers just discussed observed viewpoint-dependent recognition because their stimuli violated one or more of these conditions for viewpoint invariance. Specifically, Biederman and Gerhardstein argued that Rock et al.'s (1981) bent wire objects violate condition (1), and Tarr's cube objects and Bülthoff and Edelman's (1992) paper clip objects violate condition (2).

Biederman and his colleagues have conducted several studies in an attempt to establish that GSDs mediate object recognition and viewpoint invariance. Biederman and Cooper (1991b; see also Biederman & Cooper, 1991a, 1992) conducted repetition priming experiments in which subjects named contour-deleted line drawings of familiar objects in two blocks of trials. Drawings with complementary contours were created by deleting half of the vertices and midsegments in one drawing and deleting the opposite vertices and midsegments in another drawing. Similarly, drawings with complementary components were created by deleting half of the components in one drawing and the opposite components in another drawing. They found that name priming transferred to drawings depicting complementary contours, but not to drawings depicting complementary components. They argued that complementary-contour drawings activated the same component-based structural description, yielding repetition priming, but that complementary-component drawings activated different structural descriptions.

Biederman and Gerhardstein (1993) conducted similar experiments with intact line drawings of familiar objects to examine repetition priming following rotation in depth about the vertical axis. They found that priming was unaffected by changes of orientation as long as the same components were visible in different views. To counter arguments that familiar objects are overlearned and are thus inappropriate for studying shape representation, they conducted additional experiments with unfamiliar geon-based objects. Consistent with Biederman's (1987) recognition-by-components theory, same–different responses to sequentially presented objects were fastest and most accurate when the same object was shown from the same viewpoint; response times (RTs) were only slightly slower and accuracy only slightly worse when the viewpoint was different in the two intervals but the visible components were the same; and responses were slowest and accuracy worst when the viewpoint was different and the visible components were different.

Although these results are consistent with Biederman's (1987) claim that GSDs mediate viewpoint invariance, it should be pointed out that other researchers have failed to replicate some of these results. In particular, Srinivas (1993) found greater priming for same-contour as opposed to complementary-contour line drawings, and in another study she found that priming decreased with rotation in depth (Srinivas, 1995).

The present study attempted to account for some of these discrepancies by examining whether viewpoint dependence can be explained by the types of visual features that distinguish the objects of study. Consider what kinds of features distinguished the objects in the experiments cited. Edelman and Bülthoff's (1992) paper clip objects, for example, could be distinguished only by the 3-D connection angles between the components and the 3-D lengths of the components. The components were identically shaped and all were connected end to end. Even under the most realistic viewing conditions, these quantitative 3-D properties are difficult to recover from images. Most importantly, they may appear to be different if the object is seen from different viewpoints. Indeed, Sklar, Bülthoff, Edelman, and Basri (1993) found that recognition of paper clip objects composed entirely of cylinders suffered if the projected connection angles changed, regardless of whether the changes were due to rotation in depth or to nonrigid deformation. If features that appear different when observed from different viewpoints are the only ones that can be used to distinguish among a set of objects, one might expect recognition to be highly viewpoint dependent. A similar argument can be made about the smoothly curved wire objects studied by Rock and DiVita (1987) and the cube objects studied by Tarr (1995).

In contrast, the objects studied by Biederman and his colleagues (Biederman, 1987; Biederman & Cooper, 1991a, 1991b, 1992; Biederman & Gerhardstein, 1993)

were composed of differently shaped components that were not all connected end to end. This was true of both the familiar objects and the unfamiliar geon objects. By construction, the 3-D shapes of geons can be recovered from almost any viewpoint. Similarly, the type of connection between two components—for example, end to end or end to middle—can be recovered easily from almost any viewpoint. If objects can be distinguished on the basis of variations in these types of features, one might expect recognition to suffer less with changes of viewpoint.

The experiments presented here examine this feature-based explanation of viewpoint generalization by comparing recognition of objects that must be discriminated by features that are difficult to identify in rotated views (i.e., objects that can be distinguished only by differences in the connection angles between their components) to recognition of objects that can be discriminated by features that are relatively easy to identify in rotated views (i.e., objects that can be discriminated by the configurations of their differently shaped components). All of the objects were composed of four long, thin components connected end to end. Each component had one of three different shapes. Different objects always had different 3-D connection angles between their components. To examine generalization with viewpoint-dependent features, targets and distractors had the same component shapes connected in the same order (so that they differed only in the connection angles between the components). Although not all of the components in these objects were cylinders, as they were in Edelman and Bülthoff's (1992) experiments, the shapes of the components provided little additional information as to the identity of any particular object because all of the objects had the same components connected in the same order. To examine generalization with viewpoint-invariant features, different objects had, in addition to different 3-D connection angles, differently shaped components in at least one corresponding position in their four-component chains.

Although the three component shapes studied in these experiments were not all geons, they fulfilled an important property of geons. That is, they could be discriminated from one another from nearly any 3-D orientation. It should be made clear, however, that although the three component shapes were qualitatively distinct and could have been identified from nearly any 3-D orientation, it would have been difficult or impossible to distinguish objects with different component configurations using a representation as coarse as Biederman's GSD. In particular, there would be great overlap in the codings of the relative positions of the parts, which would not allow objects with different component configurations to be readily discriminated.

There are several benefits associated with the use of the objects studied here. First, they were unfamiliar, which allowed complete control over prior exposure and familiarity with different views of the objects. Second, because they were composed of long, thin components, occlusion of whole components was rare. Third, these

objects allowed a more strict investigation of the role of complex features, such as differently shaped components in recognition and viewpoint dependence, because these objects could not be distinguished by unique features. Tarr and Bülthoff (1995) argued that Biederman's findings can be explained by the use of unique features. In particular, they argued that the objects studied by Biederman and Gerhardstein (1993) could typically be distinguished by the presence or absence of a single highly salient feature. This was not possible in the present experiments because different objects always shared some of the same components. Finally, because these objects could not be readily distinguished by GSDs, they allow a test of whether component-based effects on viewpoint generalization must be mediated through GSDs.

EXPERIMENT 1
## Old–New Recognition

Method
Subjects. The subjects were 23 undergraduate students from the University of California, Irvine. Each received 1 h extra credit. All were naive to the purposes of the experiment and had normal or corrected-to-normal visual acuity (20/40 Snellen eye chart). The data of 3 subjects were not included in subsequent analyses because their recognition rates for unrotated objects did not exceed chance[1] levels. All 3 had participated in the different-components condition. This left 10 subjects in each components condition.

Apparatus. The stimuli were presented on a DEC 5000-200 PXG Turbo workstation with a 34.2 cm wide $\times$ 27.3 cm high display scope. The total resolution of the display scope was 1,280 pixels[2] wide $\times$ 1,024 pixels high. Two images were displayed side by side on the workstation's display scope to simulate stereoscopic viewing. Each image was 640 pixels wide $\times$ 640 pixels high. The images were viewed stereoscopically from a distance of 115 cm through a mirror stereoscope. The convergence of the eyes was set to 0°. The visual angle of each image was 8.5° on each side. The subject responded by pressing one of two buttons on a mouse.

Stimuli. The stimuli were orthographic projections of shaded 3-D objects. To simulate stereoscopic viewing, two orthographic projections of the object were presented side by side on the computer's display scope. The projection on the left depicted the object as it would appear from the 3-D location of the subject's left eye. The projection on the right depicted the object as it would appear from the 3-D location of the right eye. The angular difference between these two viewpoints was determined individually for each subject on the basis of the separation of the subject's eyes. A stereogram of one of the objects presented in this experiment is shown in Figure 1.

The computer simulation included lighting and shading information. The specular color of each object was gray (the red, green, and blue RGB values were each 0.5). The ambient, diffuse, and specular reflection components of the surface were 0.5, 0.5, and 0.3, respectively. The specular exponent was 200.0. Ambient and directional white lights were simulated to illuminate the surface. The directional light originated 34.2 cm above and 34.2 cm to the right of the simulated viewpoint. Gouraud shading was used to interpolate the surface reflectance between the object's facets. (See Gaskins, 1992, chap. 15, for a complete description of how these parameters are converted into RGB color values.)

Each object was made of four cylindrical components connected end to end. Each connection angle was chosen pseudorandomly between 25° and 155°. The object was not allowed to self-intersect. The lengths of the components were chosen pseudorandomly so that the length of the shortest component was not less than half the length
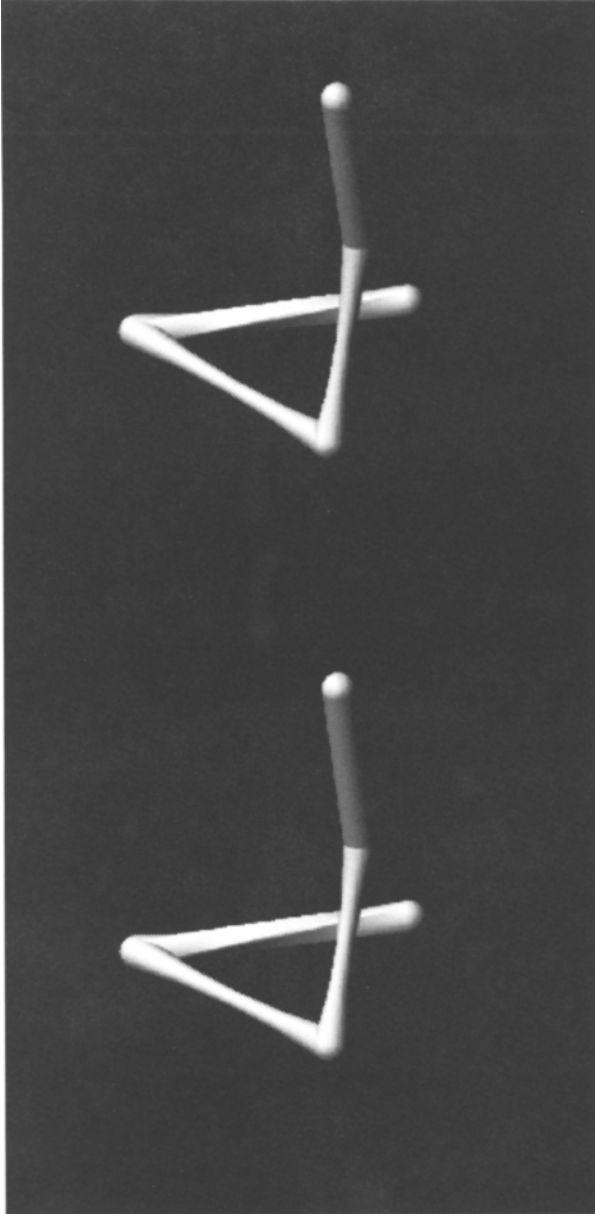
Figure 1. A stereogram of one of the objects presented in Experiment 1. After turning the page sideways, it can be viewed with a Brewster stereoscope or by diverging the eyes.

of the longest component. Each object was scaled to fit within a sphere of radius 304 pixels. Thus, the maximum possible length of a component was 608 pixels. The maximum diameter of each component was 38 pixels.

There were three differently shaped components. Each of these appears in Figure 1. All three shapes had a straight axis, but their cross sections differed. The first shape was a cylinder with a circular cross section. The second shape also had a circular cross section, but its diameter varied along the length of the axis. The diameter was maximal (the same as the cylinder) at each end and minimal in the middle. The third shape had a circular cross section at each end (equal in diameter to the cylinder) and a starlike cross section in the middle. The starlike cross section was produced by inverting the edges in the four quadrants of a circle. The cross section of the com-

ponent varied smoothly along the length of the component from a circle to a star and then back to a circle. The four points of the star produced an edge on the component's surface. To make this edge more visible, the component was twisted 90° along the length of the component.

Since all of the objects consisted of four components connected end to end, it was possible to code each object by the shapes of the components occupying each of its four positions. Two objects were considered qualitatively different if the shapes occupying at least one corresponding position were different. Two qualitatively different objects could have the same components in different orders. An object coded cyl:cyl:star:cons, for example, was qualitatively different from an object coded cyl:star:cyl:cons, even though they contained the same number of each component shape. An object coded cyl:cyl:star:cons was equivalent to an object coded cons:star:cyl:cyl, however, since the order of the components had to be different relative to either end of the four-segment chain for two objects to be qualitatively different.

Every object contained either two or three differently shaped components. These restrictions yielded 42 equivalence classes of objects. An equivalence class was defined as a set of objects having the same components connected in the same order. The connection angles between the components and the lengths of the components, however, were different for different members of the class. Objects from the same equivalence class can be said to differ quantitatively; objects from different equivalence classes can be said to differ qualitatively. Examples of objects from different equivalence classes are shown in the top row of Figure 2. Examples of objects from the same equivalence class are shown in the bottom row of Figure 2. Only the lengths of the components and the connection angles between the components were different for objects from the same equivalence class.

Procedure. Subjects were run individually in single 40-min sessions. The session was divided into nine blocks of trials. The first block was practice, and the data from it were not included in the analyses. At the beginning of each block, the subject viewed six target objects. The objects were presented one at a time for 7 sec each. Each object was preceded by a 750-msec cross to aid stereo convergence. After a 30-sec break, the subject viewed 12 objects. Half of these were the same target objects and the remainder were distractor objects. The presentation order of the six targets and six distractors was completely randomized. Each object was preceded by a 750-msec cross, and each object remained visible until a response was made. The subject's task was to decide whether each of the 12 objects seen after the 30-sec break had been seen before the break. The subject was informed that the orientation of the targets could change. The subject pressed the left button on the mouse to indicate that the object had been seen before the break and pressed the right button on the mouse to indicate that it had not been seen. The computer beeped if the response was incorrect.

Design. There were two independent variables. The first, coding of components, was run between subjects. For half of the subjects, all 12 objects (6 targets and 6 distractors) in an experimental block came from the same equivalence class. Thus, there were only quantitative differences (different component lengths and connection angles) between the targets and distractors. For each block, the equivalence class was selected at random from the 42 classes. These subjects were termed the "same-components" group because all of the objects viewed by them in an experimental block had the same components connected in the same order.

For the remaining subjects, each of the 12 objects (6 targets and 6 distractors) in an experimental block came from a different equivalence class. Thus, there were qualitative as well as quantitative differences between any 2 of the 12 objects. Because no two objects in an experimental block had the same components in the same order, these subjects were termed the "different-components" group. For each experimental block, 12 equivalence classes were chosen at
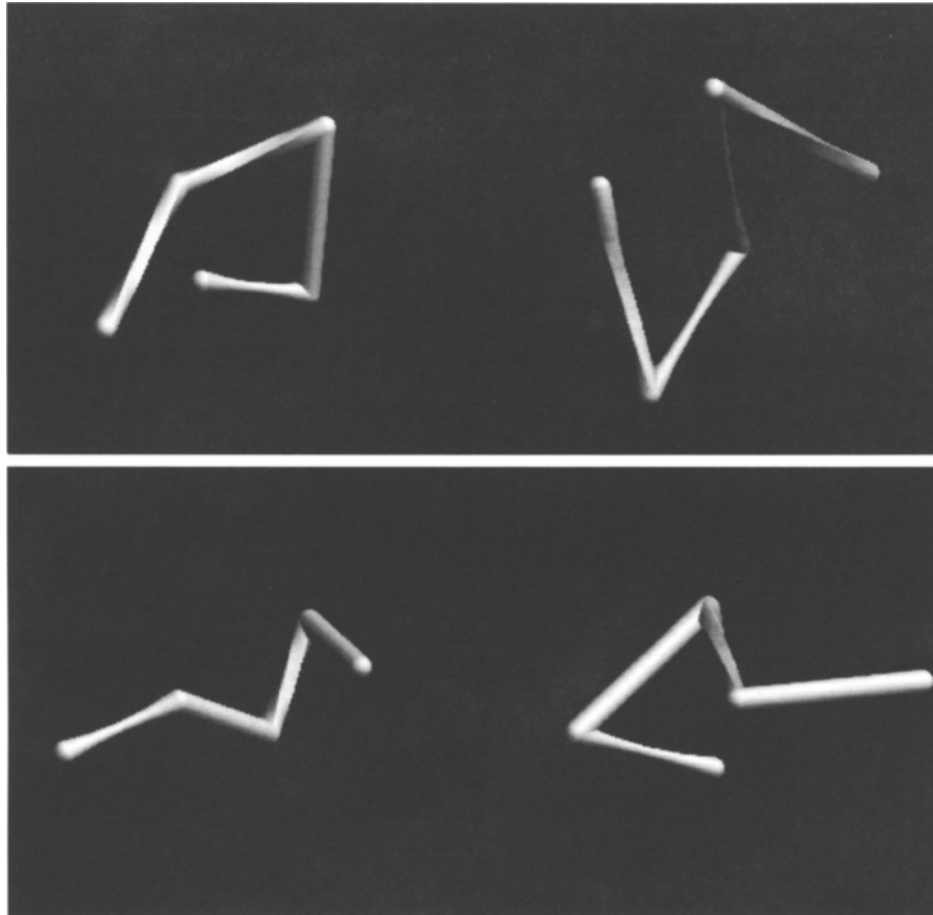
Figure 2. Four objects presented in Experiment 1. The objects in the top row are qualitatively different because the components occupying each place in the four-segment chain are not all the same. The objects in the bottom row are quantitatively different because the components occupying each place in the four-segment chain are the same in both objects. Only the lengths of the components and the connection angles between the components are different.

random from the 42 classes. Half of these were randomly selected to be the targets, and the other half served as distractors. New targets and distractors were used for each block and for each subject.

The second independent variable, which was run within subjects, was the orientation of the target objects in the recognition phase of each block. Two of the target objects were presented in their original orientation; two were rotated 22.5° about the vertical axis from their original orientation, and two were rotated 45° about the vertical axis from their original orientation. The direction of rotation (clockwise or counterclockwise) was determined randomly for each rotated object. Which of the six targets were rotated was also determined randomly.
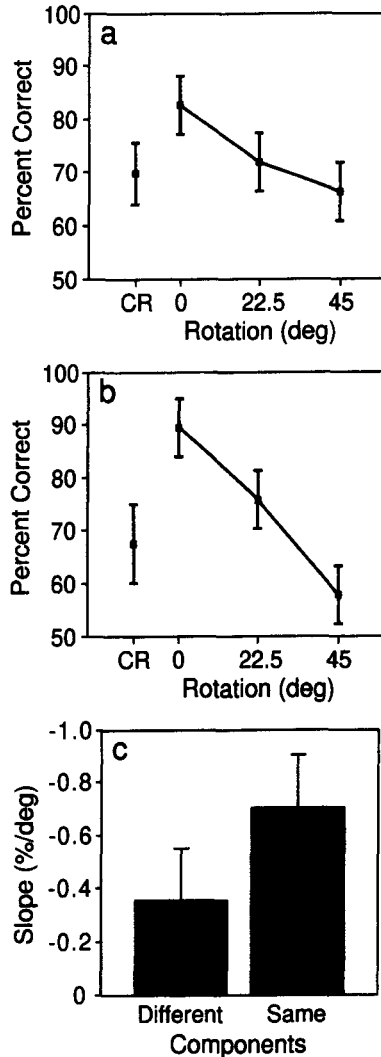
Results

The percent of target objects correctly recognized (the hit rate) was computed for each subject at each of the three levels of target orientation. Each hit rate was based on 16 signal trials. A single correct rejection rate was computed for each subject on the basis of 48 distractor trials. The mean correct rejection rates and the hit rates for each level of target orientation (0°, 22.5°, and 45°) are pre-

sented in Figure 3. Panel a shows the results for subjects in the different-components condition, and panel b shows results for subjects in the same-components condition. The error bars indicate 95% confidence intervals. Within-subjects confidence intervals (see Loftus & Masson, 1994) were computed separately over target orientation for subjects in the two conditions. It is clear from the figure that recognition declined in both conditions when the target objects were rotated from the studied view. As shown in panel c, recognition declined at a faster rate with changes in orientation for subjects in the same-components condition.

Discussion

This experiment demonstrates that viewpoint generalization varies with the type of features that distinguish the objects. Recognition of objects distinguishable by the configurations of their components declined less with rotation from the studied view than did recognition of

Figure 3. The percent of correct recognition judgments in Experiment 1. Results for subjects in the different-components condition are shown in panel a. Results for subjects in the same-components condition are shown in panel b. Panel c shows the best fitting slopes for the two conditions. Error bars indicate 95% confidence intervals. CR, correct rejections.

objects distinguishable only by the lengths of their components and the connection angles between their components. The value of assessing viewpoint dependence for both classes of objects is evident since recognition was not completely viewpoint invariant for either class. If only the different-component objects had been studied in this experiment, one might have concluded that variations in component configuration have no influence on the magnitude of viewpoint dependence. Comparing the influence of component variations with variations in connection angles, however, has revealed that the variations in component configuration studied here reliably facilitate recognition of novel views of objects, although they do not necessarily confer complete viewpoint invariance.

Several explanations of the differences in viewpoint dependence observed here can be dismissed easily. Some researchers have argued that recognition is sometimes less viewpoint dependent because multiple views of the objects are stored following repeated exposure to the objects (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992; Tarr, 1995; Tarr & Pinker, 1989). Since the objects studied here were unfamiliar and were studied from only a single viewpoint, differences in viewpoint dependence cannot be attributed to differences in familiarity or differences in experience with different views of the objects. The present results do not necessarily argue against the use of multiple views to achieve viewpoint invariance. Rather, they suggest that the features available to distinguish among a set of objects can significantly influence the range of viewpoints over which a single studied view is effective for recognition.

Biederman and Gerhardstein (1993) argued that viewpoint dependence can often be attributed to the accretion and deletion of visible components that results from rotation in depth. They argued that changes in the visible components alter the GSD that is derived from the image, hindering recognition of objects in rotated views. The objects studied in the present experiments were constructed of long, thin components that made accretion and deletion of whole components rare. Furthermore, recognition of rotated objects was enhanced, not hindered, in the different-components condition. It is in this condition that accretion and deletion of components would have had the greatest potential to influence recognition. In the same-components condition, changes in the visibility of particular components would not have been informative because all of the objects in an experimental block had the same components connected in the same order.

Finally, in some previous experiments in which viewpoint-invariant recognition was observed, different objects in the stimulus set could be distinguished by unique, orientation-invariant features (e.g., Eley, 1982). This was true of the novel objects studied in Biederman and Gerhardstein's (1993) Experiments 3 and 5, in which each object contained a unique, differently shaped component. Tarr and Bülthoff (1995) argued that Biederman and Gerhardstein's explanation of their viewpoint-invariant results in terms of GSDs cannot be differentiated from an explanation based on unique features. They further argued that although such unique features might support viewpoint-invariant recognition for a restricted set of objects, they are unlikely to be responsible for viewpoint-invariant recognition in natural viewing conditions because such features are unlikely to be diagnostic within the larger set of objects encountered in natural viewing. The different-component objects studied in the present experiments could not be distinguished on the basis of unique features. That these objects were more easily recognized from novel viewpoints suggests that unique features are not necessary to facilitate viewpoint-invariant recognition. Recognition of these objects was not completely viewpoint invariant, however, leaving open the

possibility that unique features are necessary to achieve complete viewpoint invariance following exposure to only a single view of an object.

An unexpected finding in this experiment was that subjects in the different-components condition recognized fewer unrotated objects than did subjects in the same-components condition (although overall percent correct was similar for both groups). One might have expected those subjects viewing different-component objects to perform better at all orientations because they could have recognized target objects on the basis of their component configurations and on the basis of other features, such as projected connection angles, that were available to both groups. One possible explanation of this finding is that in using the component configurations, subjects in the different-components condition attended less to features such as the projected connection angles and thus recognized fewer unrotated objects relative to subjects in the same-components condition, who could rely only on these features. Features such as the projected connection angles are most useful for recognizing unrotated objects because they change considerably if the object is rotated in depth from the studied viewpoint. If subjects in the different-components condition did not encode these features as well as subjects in the same-components condition, then they would have recognized fewer unrotated objects.

In their experiments comparing viewpoint dependence for bent wire and surface objects, Farah, Rochlin, and Klein (1994) also found a crossover interaction, such as the one shown in Figure 3. They hypothesized that even though recognition of surface objects suffered less with changes of view than did recognition of wire objects, surface objects were overall more difficult to recognize. This hypothesis was supported by their correct rejection rates—new wire objects were more often identified as new than were new surface objects. This was not the case in the present experiment. Overall percent correct was roughly equivalent for both groups of subjects. This suggests that different-component objects were not simply more difficult to recognize. To address this issue directly, a control experiment was conducted. To concentrate attention on the unrotated objects, none of the target objects presented during the test phase of an experimental block were rotated with respect to their original orientation (i.e., all target objects were tested at 0°).

Control Experiment

Twenty-six undergraduate volunteers from the University of California, Irvine, participated in the control experiment. Each received 1 h of extra credit, and all were naive to the purposes of the experiment. The data of 6 subjects were not included in the analyses because their recognition rates did not exceed chance levels. Four had participated in the different-components condition, and 2 had participated in the same-components condition. This left 10 subjects in each group.

The percent of hits and correct rejections was computed for each subject based on 48 signal and 48 distractor trials. Neither the hit rates nor the correct rejection rates were reliably different. The mean hit rates were 83.8 and 84.0 and the mean correct rejection rates were 74.4 and 78.8 for subjects in the same-components and different-components conditions, respectively.

The difference in the hit rates for unrotated objects that was observed in the main experiment was not replicated in the control experiment, indicating that objects with different component configurations are not simply more difficult to recognize. Poorer recognition of unrotated objects with different component configurations is contingent on the presence of rotated target objects during the testing phase of the experiment. This suggests that the influence of different features on recognition depends not only on which features are available to distinguish the objects, but also on their relevance to performing the task. In the control experiment, it would have been possible to recognize target objects entirely on the basis of their image appearance: Novel images never corresponded to targets. In the main experiment, it would not have been possible to recognize target objects solely on the basis of their projected images since two thirds of the target objects were rotated from the studied view during testing. This would have prompted subjects to seek out other features that were recognizable in rotated views. The variations in component configurations could easily have served this purpose for subjects in the different-components group.

The results of the control experiment suggest that the influence of different features on recognition is not absolute, but instead depends on factors such as the relevance of the features to performing the task. To examine this issue more directly, subjects in Experiment 2 were biased while studying the objects to attend to either the connection angles between the components or the shapes of the components themselves. Both groups of subjects viewed objects that could be distinguished on the basis of their component configurations. If, as in Experiment 1, making use of one type of feature compromises the use of other features, then attending to the components rather than to the connection angles could hinder recognition of unrotated objects as well as enhance recognition of rotated objects.

EXPERIMENT 2
Attention to Features

The results of Experiment 1 indicate that recognition of objects with different configurations of components suffers less with changes of viewpoint than does recognition of objects that differ primarily in the connection angles between their components. It was argued that this difference occurred because the shapes of the components in a particular object are easier to identify in rotated views than are the connection angles between components.

The present experiment examined the components-versus-angles interpretation more closely by determining whether specifically attending to one or the other of these features influenced viewpoint dependence, as did the availability of these features in Experiment 1. All of the subjects viewed objects with different component configurations. Thus, regardless of which features were actually attended, the component-based information that was found to enhance viewpoint invariance in Experiment 1 was, in principle, available. Attending to the connection angles between an object's components might disrupt the subject's ability to use the information that would normally be conveyed by variations in the configurations of the components. This might yield viewpoint-dependent recognition performance because, as shown in Experiment 1, the connection angles do not readily support recognition following rotation in depth.

This was a surprise recognition experiment. Subjects did not know while studying the objects that they would have to recognize them later in the experiment. Different tasks were given to the subjects during the study phase of the experiment. Half of the subjects (the components group) were instructed to report the number of differently shaped components that were present in each object. This was intended to bias the subjects to attend to the component structure of the objects. The remaining subjects (the angles group) were instructed to report the number of connection angles in each object that were less than 90°.

Method

Subjects. The subjects were 46 undergraduate students from the University of California, Irvine. Each received ½ h of extra credit for participating. All were naive to the purposes of the experiment and had normal or corrected-to-normal visual acuity. There were 23 subjects in each study condition.

Stimuli. The stimuli were constructed as they were in Experiment 1. All of the subjects viewed objects with different configurations of components.

Procedure. Subjects were run individually in 15-min experimental sessions. The session consisted of two parts. In the first part, the subject was shown six objects, one at a time, for 7 sec each. After viewing each object, subjects in the components group were instructed to report the number of differently shaped components that were present in the object. Subjects in the angles group were instructed to report the number of connections between components that were less than 90°. (No mention was made as to whether the subject should report the 2-D projected angle or the 3-D angle. None of the subjects questioned this instruction. Debriefing indicated that most subjects had attempted to report on the 3-D angles.) These study tasks were selected because each requires that the subject look at every component in the object at least once.

After the subject had responded to all six objects, the experimenter read the instructions for the second part of the experiment. The subject was instructed to decide, for each object to be presented, whether it had been seen in the first part of the experiment. The subject was informed that the viewpoint from which the objects had been seen could change. The subject responded verbally with "old" or "new." In addition to the old–new decision, the subject was instructed to provide aloud a confidence rating for the decision. A rating of "one" meant that the subject was not confident in the decision; a rating of "two" meant that the subject was some-

what confident; and a rating of "three" meant that the subject was very confident in the decision. Each subject repeated this procedure for 12 objects, 6 of which were the objects seen in the first part of the experiment, and 6 of which were distractor objects.

Only one block of data was collected for each subject because, after completing one block, the subject knew that the objects would have to be recognized in the second part of the experiment. This might have prompted the subject to adopt a new strategy while studying the objects in later blocks.

Design. There were two independent variables in this experiment: (1) the study instruction given to the subject ("Report the number of differently shaped components" or "Report the number of connection angles less than 90°"), and (2) the orientation of the target objects during the recognition phase of the experiment. Two of the target objects were presented in the same orientation as in the study phase of the experiment; two were rotated 22.5° about the ver-
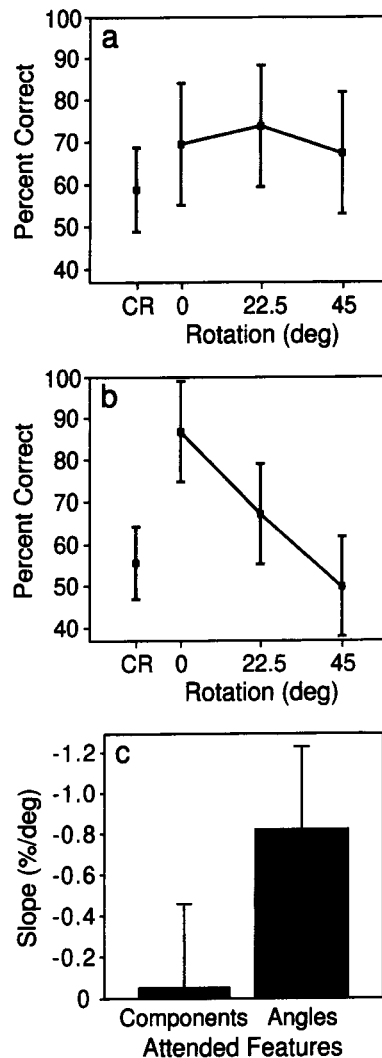

Figure 4. The percent of correct recognition judgments in Experiment 2. Results for subjects who reported the number of differently shaped components during the encoding phase are shown in panel a. Results for subjects who reported the number of angles less than 90° are shown in panel b. Panel c shows the best fitting slopes for the two conditions. Error bars indicate 95% confidence intervals. CR, correct rejections.

tical axis, and two were rotated 45° about the vertical axis. Encoding instruction was run between subjects. The subjects were alternately assigned to the study conditions as they arrived. Target orientation was run within subjects. As noted, all 12 of the objects shown to each subject (6 target objects and 6 distractor objects) had different component configurations (as in the different-components condition of Experiment 1).

Results

Each subject could recognize zero, one, or two objects at each of the three target orientations. Correct rejection rates were based on six distractor objects. The percent of correct rejections and the percent of targets recognized at each orientation (i.e., the hit rates) are presented in Figure 4. Panel a shows the results for the subjects who reported the number of differently shaped components in each object during the encoding phase of the experiment. Panel b shows results for the subjects who reported the number of connection angles in each object that were less than 90°. Since there were fewer responses collected for each subject in this experiment than in Experiment 1, error bars showing 95% confidence intervals are somewhat larger. Nevertheless, there are clear differences in performance for the two groups of subjects in this experiment. There was no clear decline in performance with rotation from the studied view for subjects reporting on the components. However, there was a substantial decline in recognition for subjects reporting on the angles. The difference in performance between the two conditions is demonstrated in panel c, which shows the slopes of the functions in panels a and b.

The confidence ratings revealed that subjects in the components group were consistently more confident in their recognition decisions than were subjects in the angles group. Although their confidence for correct responses declined somewhat with rotation from the studied viewpoint, they were still more confident in their recognition decisions for objects rotated 45° than were subjects in the angles group for any viewpoint. For both groups, confidence ratings were consistently higher for correct responses.

Discussion

Recognition by subjects who had reported the number of differently shaped components was unaffected by rotation from the studied viewpoint. On the contrary, recognition declined sharply with change of viewpoint for the subjects who had reported the number of connection angles less than 90°. These results support the Experiment 1 conclusion that differences in viewpoint dependence arose because subjects in different conditions made use of different visual features, some more stable over changes of viewpoint than others.

This experiment further indicates that the presence of features capable of supporting viewpoint-invariant recognition is not necessarily sufficient to yield viewpoint-invariant performance. Although all of the subjects could have used the variations in component configurations to recognize rotated objects, only those who had been in-

structed to report the number of differently shaped components appear to have benefited from these variations. The present experiment demonstrates that characteristics of the task, including the instructions to the subject, can reliably affect viewpoint dependence. This finding might help to explain some of the inconsistencies in viewpoint dependence observed by other researchers. For example, using the same set of unfamiliar objects, Tarr, Hayward, Gauthier, and Williams (1994) observed viewpoint-dependent recognition, whereas Biederman and Gerhardstein (1993) observed viewpoint-invariant recognition. Since the RTs reported by these researchers were very dissimilar, it is possible that some aspects of their method were sufficiently different to induce the subjects to use different features.

GENERAL DISCUSSION

These experiments examined the hypothesis that viewpoint dependence is determined by the types of features that distinguish the objects of study. Specifically, it was expected that recognition of objects distinguishable by features that can be recovered from only a limited range of viewpoints would be best for studied viewpoints and would decline with rotation in depth from the studied viewpoints. In contrast, it was expected that recognition of objects distinguishable by features that can be recovered from a wider range of viewpoints would suffer less with changes of viewpoint. This hypothesis was confirmed in Experiment 1. Recognition of objects with distinct component configurations declined less with changes of viewpoint than did recognition of objects with identical component configurations varying primarily in the connection angles between their components. The 3-D connection angles are difficult to recover from images and may appear different from different viewpoints. These features do not support viewpoint-invariant recognition. Objects with different component configurations could be distinguished by the absolute and relative image locations of their differently shaped components. These features allow enhanced view generalization because the shapes of the components can be recovered from a wide range of viewpoints. Experiment 2 indicated that the presence of such features is not necessarily sufficient to produce viewpoint-invariant recognition performance. Subjects biased to attend to viewpoint-specific features exhibited viewpoint-dependent recognition, although features capable of supporting viewpoint-invariant performance were present in the stimuli.

Takano (1989) presented an interesting discussion of how the types of features that are used in a recognition task can affect viewpoint dependence. He distinguished orientation-free and orientation-bound information as well as elementary and conjunctive relational information. He argued that a transformation mechanism such as mental rotation is needed to make use of orientation-bound conjunctive features, which would yield viewpoint-dependent recognition performance, but that no such

mechanism is necessary to make use of orientation-free information, whether elementary (specifying, e.g., whether a contour is straight or curved) or conjunctive (specifying, e.g., whether two contours are connected or not). Regarding the present experiments, it is possible that the different-component objects contain more orientation-free features, allowing them to be recognized without the need for transformation mechanisms. Precisely what those features might be is, at present, unclear. Nevertheless, the results of Experiment 2 suggest that they are likely related to the features that distinguish the different part shapes. As noted, it would have been difficult or impossible to discriminate between targets and distractors using single, unique features, so recognition of the different-component objects might have required the use of some orientation-bound conjunctive features. According to Takano's theory, this would have led to at least some viewpoint dependence, as was observed in Experiment 1.

Tarr and Pinker (1990) provided a somewhat more specific argument along the lines of orientation-free and orientation-bound conjunctive information. They argued that mental rotation is not necessary to recognize disoriented objects if the spatial arrangements of their parts can be distinguished along a single dimension. In contrast, they argued that mental rotation is necessary if more than one dimension is needed to distinguish the arrangements.

Jolicoeur (1990) proposed a dual-system theory to explain why recognition is sometimes viewpoint invariant and sometimes viewpoint dependent. Jolicoeur postulated that that two systems—one feature based and orientation invariant, the other wholistic and orientation dependent—function independently and in parallel, with recognition time determined by the system that finds a match to memory first. Jolicoeur argued that the orientation-dependent system must rely on mental rotation to transform disoriented patterns to the nearest orientation stored in memory. The feature-based system, in contrast, can function without mental rotation. Although Jolicoeur's theory was developed to explain differences resulting from rotations in the image plane, it would be natural to extend it to rotation in depth. With regard to the present experiments, it is possible that the different-component objects contained features that allowed recognition via the viewpoint-invariant, feature-based system, whereas the same-component objects lacked such features, so recognition of these objects would have been possible only by using the viewpoint-dependent system.

The results of the present study are in close agreement with those of a recent study by Tarr, Bülthoff, Zabinski, and Blanz (1997). They conducted a sequential matching task with novel objects composed of five geons connected end to end. In the no-parts condition, both objects were composed entirely of cylinders, so different objects had to be discriminated by the connection angles between the cylinders. On *same* trials, RTs increased and accuracy decreased with greater angular disparity between the two objects. In the three- and five-part condi-

tions, different objects had (in addition to different connection angles) different components in three or five of the positions within the five-segment chains, respectively. As in the present study, recognition was still dependent on the angular disparity, but recognition did not decline as rapidly with greater angular disparity. Viewpoint-invariant performance was not even attained in the one-part condition, in which different objects could be discriminated by the shape of a single component. Tarr et al. obtained similar results in an identification task in which subjects learned names for a small number of their five-segment chain stimuli.

Two other recent studies bear on the discussion of features presented here. Farah et al. (1994) examined whether recognition of unfamiliar "surface" objects made of bent clay disks was more viewpoint invariant than recognition of bent wire objects like those studied by Rock and his colleagues (Rock & DiVita, 1987; Rock et al., 1981), which were made by tracing the edge of each disk with a wax-covered string. They found that recognition of the surface objects was less dependent on viewpoint than was recognition of the wire objects, suggesting that features associated with a surface-based description of object shape facilitate viewpoint invariance. Liu, Kersten, and Knill (in press) studied discrimination of objects composed of collections of disconnected balls, the same balls connected by thin line segments, and the same balls connected by thick lines (so that they became four-segment "paper clips" composed entirely of cylinders). Thresholds for discriminating these objects from objects of the same type with small displacements of the balls were greatest for isolated balls, less for balls connected by thin lines, and least for paper clips. These results are consistent with the feature-based explanation of viewpoint generalization in that the paper clip objects are more likely to contain surface features that better specify their shapes.

The differences in viewpoint dependence observed in the present study cannot be explained easily by current part-based structural description theories or by current multiple-views theories. Biederman's (1987) recognition-by-components theory predicts no difference in the recognition of same- and different-component objects, because GSDs do not adequately represent spatial relations among parts to discriminate either the same- or different-component objects studied here. Multiple-views theories (Bülthoff & Edelman, 1992; Tarr, 1995) generally have not considered the importance of the types of features that are available to distinguish the objects.

Overall, these results indicate that a more comprehensive, feature-based model of shape representation must be developed to account for differences in viewpoint dependence. The present experiments indicate that the model must represent a variety of visual features, and it must be able to adapt to changing visual environments, perhaps by learning to rely on different features. Experiment 2 demonstrates that the influence of various features on recognition performance is not absolute. Even though

component variations were present, attending to the connection angles between the components led to viewpoint-dependent performance.

Edelman (1991; see also Bülthoff & Edelman, 1993) outlined one such model of shape representation that begins with the extraction of a variety of different features, ranging from variations in local image intensity to components defined by nonaccidental properties such as colinearity and axis curvature. A vector specifying the image locations of these features is then processed in a view-interpolation classification network—for example, one utilizing hyper basis functions (Poggio & Edelman, 1990; Poggio & Girosi, 1990). The system's ability to use a variety of features makes it extremely versatile. Edelman (1991) argued that such a system should be able to adapt itself to rely on different features depending on the demands of the task. Edelman also pointed out that there could be a tradeoff between the diagnosticity of a particular feature and its ability to support viewpoint-invariant recognition. To take an example from the present study, features such as connection angles are extremely diagnostic because small variations in them can be used to discriminate among very similar objects. Since these features are difficult to recover from images and might appear different in different views of the same object, they do not support viewpoint-invariant recognition. Features such as differently shaped components are less diagnostic because many different objects are likely to have roughly the same components arranged in much the same way. However, these features are relatively easy to recover from different viewpoints, making recognition of objects that *are* distinguishable by variations in these features more invariant with changes of viewpoint.

An important advantage of a recognition system that can adapt itself to rely on different visual features in different circumstances is that separate recognition systems are not necessary to accomplish basic- and subordinate-level classification (Edelman, 1991). Rather, differences in viewpoint dependence for basic- and subordinate-level classification can be explained easily by the types of features that are naturally available to distinguish objects between and within classes. The constituent components of objects in different basic-level classes typically differ (Tversky & Hemenway, 1984). Since these features support viewpoint-invariant recognition, basic-level classification should be relatively viewpoint invariant. On the contrary, objects in the same basic-level class are likely to share the same components. Thus, to discriminate objects in the same basic-level class, the visual system must rely on features that are less able to support viewpoint-invariant recognition.

Biederman and his colleagues have argued that some viewpoint-dependent effects in recognition are due to the influence of visual processing not related to recognition—for example, processing within a system for visual–motor interaction (Biederman & Cooper, 1991a, 1992). An explanation of viewpoint-dependence relying on multiple vi-

sual representations is not necessary if the visual system is capable of using differently shaped features. According to the feature-based approach advanced here, viewpoint-dependent recognition is not due to the influence of non-recognition visual systems, but instead reflects a reliance on viewpoint-specific features. Why and under what circumstances the visual system might rely on such features is an empirical question requiring further research.

The present study provides some insights into this issue. Experiment 1 demonstrates that viewpoint dependence is related to the features that distinguish the objects. If these features cannot be recovered in rotated views, recognition will suffer with changes of view. Experiment 2 demonstrates that viewpoint dependence is also related to the task. If the task requires subjects to attend to viewpoint-specific features, recognition can be viewpoint dependent, even if viewpoint-invariant features are available to distinguish the objects.

REFERENCES

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

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

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

BIEDERMAN, I., & COOPER, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception & Performance*, 18, 121-133.

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

BÜLTHOFF, H. H., & EDELMAN, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*, 89, 60-64.

BÜLTHOFF, H. H., & EDELMAN, S. (1993). Evaluating object recognition theories by computer graphics psychophysics. In T. A. Poggio & D. A. Glaser (Eds.), *Exploring brain functions: Models in neuroscience* (pp. 139-164). New York: Wiley.

EDELMAN, S. (1991). *Features of recognition* (Tech. Rep. fCS-TR10). Rehovot, Israel: Weizmann Institute of Science.

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.

ELEY, M. G. (1982). Identifying rotated letter-like symbols. *Memory & Cognition*, 10, 25-32.

FARAH, M. J., ROCHLIN, R., & KLEIN, K. L. (1994). Orientation invariance and geometric primitives in shape recognition. *Cognitive Science*, 18, 325-344.

GASKINS, T. (1992). *PHIGS programming manual*. Sebastopol, CA: O'Reilly & Associates.

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

JOLICOEUR, P. (1990). Identification of disoriented objects: A dual-systems theory. *Mind & Language*, 5, 387-410.

LIU, Z., KERSTEN, D., & KNILL, D. C. (in press). Dissociating stimulus information from internal representation—a case study in object recognition. *Vision Research*.

LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1, 476-490.

MARASCUILO, L. A. (1970). Extensions of the significance test for one-

parameter signal detection hypotheses. *Psychometrika*, **35**, 237-243.

MARR, D., & NISHIHARA, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B*, **200**, 269-294.

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

POGGIO, T., & GIROSI, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, **247**, 978-982.

ROCK, I., & DIVITA, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, **19**, 280-293.

ROCK, I., DIVITA, J., & BARBEITO, R. (1981). The effect on form perception of change of orientation in the third dimension. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 719-732.

SHEPARD, R. N., & COOPER, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.

SKLAR, E., BÜLTHOFF, H. H., EDELMAN, S., & BASRI, R. (1993). Generalization of object recognition across stimulus rotation and deformation. *Investigative Ophthalmology & Visual Science*, **34**(4, Suppl.), 1081.

SRINIVAS, K. (1993). Perceptual specificity in nonverbal priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 582-602.

SRINIVAS, K. (1995). Representation of rotated objects in explicit and implicit memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1019-1036.

TAKANO, Y. (1989). Perception of rotated forms: A theory of information types. *Cognitive Psychology*, **21**, 1-59.

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: Human Perception & Performance*, **21**, 1494-1505.

TARR, M. J., BÜLTHOFF, H. H., ZABINSKI, M., & BLANZ, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science*, **8**, 282-289.

TARR, M. J., HAYWARD, W. G., GAUTHIER, I., & WILLIAMS, P. (1994, November). *Geon recognition is viewpoint dependent*. Paper presented at the annual meeting of the Psychonomic Society, St. Louis.

TARR, M. J., & PINKER, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, **21**, 233-282.

TARR, M. J., & PINKER, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science*, **1**, 253-256.

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

NOTES

1. Chance performance was determined by computing d'. The significance of d' was assessed using Marascuilo's (1970) z test with p set to .05

2. Pixels were also used to measure lengths in the 3-D model used to generate the displays.