Journal of Experimental Psychology: Human Perception and Performance
In press

Effects of outline shape in object recognition

William G. Hayward

University of Wollongong

Running Head: Outline shape

Abstract

The use of outline shape in recognizing objects was investigated in four experiments. In Experiment 1, subjects matched a shaded image to either another shaded image or a silhouette. In Experiment 2, subjects initially named shaded images; later they named either shaded images or silhouettes. Performance in both experiments was predicted by changes in the outline shape of the stimuli. The same matching (Experiment 3) and priming (Experiment 4) paradigms were then used to investigate recognition with objects that were rotated between presentations so as to change the outline shape of the object. Recognition was predicted by changes to outline shape. These results place constraints on models of object recognition, and are most compatible with viewpoint-dependent models of recognition.

Because we are mobile organisms and live in a three-dimensional world, the ability to compensate for changes in viewpoint is a crucial aspect of our ability to recognize objects. When the observed viewpoint of an object changes, we usually judge that the new visual stimulus corresponds with the previously observed object, rather than corresponding with an entirely new object. Recently, a number of different researchers have proposed competing explanations for visual compensation across changes in viewpoint. These theories are typically conceived as being divided into viewpoint-invariant and viewpoint-dependent classes (e.g., Biederman & Gerhardstein, 1995; Hummel, 1994a; Tarr & Bülthoff, 1995). A theory is said to be viewpoint invariant if it predicts no cost to recognition across a rotation in depth; theories that propose costs to recognition with changes in viewpoint are considered to be viewpoint dependent. Examples of these theories are discussed below. In many ways, however, the dichotomy presented between viewpoint-invariant and viewpoint-dependent models of recognition is a false one. Rather than being two qualitative categories, these theories can be considered as different points of a single continuum. In general, viewpoint-invariant theories propose that recognition does depend on viewpoint to some extent, because single representations normally encode only some viewpoints of an object; a number of representations may be needed to cover all possible views of the object. Similarly, viewpoint-dependent theories of recognition do not necessarily propose that all changes in viewpoint will lead to a cost in recognition performance. There may be some changes to the stimulus, such as small rotations or mirror-reflections, that can be captured with a viewpoint-specific representation. Thus, the issues addressed by object recognition experiments are not so much whether a change in viewpoint will affect recognition performance, but rather in which situations a cost to recognition will occur.

Of particular interest in this paper is whether costs to recognition are predicted by changes in the outline shape of an object. The purpose of this paper is not to adopt any particular theoretical position, but rather show constraints that must be placed upon any account of object recognition across viewpoint change. In particular, the experiments reported here will show that changes in the outline shape of a stimulus are closely related to costs in the recognition of that stimulus. Thus, any theory of object recognition must be able to explain how objects can be recognized using only outline shape information.

Theories of object recognition across viewpoint change

Biederman and his colleagues have proposed that recognition of an object will be affected by changes in the visible parts of the object. Drawing on volumetric approaches to object representation (e.g., Binford, 1971; Marr & Nishihara, 1978), Biederman and his colleagues (Biederman, 1987; Biederman & Gerhardstein, 1995; Hummel & Biederman, 1992) have proposed that perceptual processes derive the constituent parts of an object, and then represent each of those parts as a simple geometric volume, or geon. An object representation, or geon structural description, consists of geons corresponding to the two or three most salient parts of an object, and the spatial configuration in which the geons are connected. A structural description is represented without regard for the specific viewpoint of the observer. Thus, whenever the same object is viewed with the same visible parts in the same spatial arrangement, an identical structural description will be derived, producing recognition performance that is invariant over changes in viewpoint. A number of studies have shown that costs to recognition are predicted by changes in the visible parts of an object; these studies have changed either the information available from a particular viewpoint of an object (e.g., Biederman & Cooper, 1991a, 1991b, 1991c, 1992; Biederman & Ju, 1988; Cooper & Biederman, 1993; Cooper, Biederman, & Hummel, 1992), or the observed viewpoint of an object (Biederman & Gerhardstein, 1993).

Other researchers have suggested that object representations are much more specific to the viewpoint of the object than simply the visible parts. For example, Tarr and colleagues (Hayward & Tarr, in press; Tarr, 1995; Tarr, Bülthoff, Zabinski, & Blanz, in press; Tarr, Hayward, Gauthier, & Williams, 1994) have shown that with a variety of objects, recognition latencies increase with larger rotations in depth between probed viewpoints, regardless of whether parts are occluded by the rotation. Such effects are attributed to “normalization”, an analog process of transforming the perceived and/or remembered images until they are able to be directly matched. Similar findings for recognition of depth-rotated objects have been reported by Edelman and colleagues (Bülthoff & Edelman, 1992; Cutzu & Edelman, 1994; Edelman & Bülthoff, 1992; Vetter, Poggio, & Edelman, 1994), Srinivas (1993), Lawson & Humphreys (1996) and Rock and DiVita (1987). These results also correspond to those for objects misoriented in the picture plane, where linear effects have been found over wide ranges of orientation change in both judgments of recognition (Jolicoeur, 1985; Tarr & Pinker, 1989) and whether a figure is identical or mirror-reversed relative to another (Cooper & Shepard, 1973; Shepard & Cooper, 1982).

There is some disagreement as to whether parts or other types of information are the basis for depth-rotated object recognition in “normal” circumstances (see Tarr & Bülthoff, 1995, and Biederman & Gerhardstein, 1995, for a summary of the debate). Regardless of the outcome of this debate, studying object recognition across depth rotation is not the only possible methodology for examining the information used in object recognition decisions. As noted above, other experiments, primarily those of Biederman and colleagues, have examined the effects on recognition of a change in visible information of a stimulus when the observed viewpoint remains identical between presentations. For example, Biederman and Cooper (1991a) examined name priming with line drawings of objects. They created two versions of each object, with each version showing half the line segments, but all the parts, of the object. Subjects named one contour-deleted version of the object, and then either the same version or the complementary version. Both showed equivalent levels of priming, leading the authors to conclude that priming was mediated by a part-based, rather than contour-based, representation. Biederman and his colleagues have examined the effects of changes in size (Biederman & Cooper, 1992; Fiser & Biederman, 1995), spatial position (Cooper, Biederman, & Hummel, 1992), hemifield of presentation (Biederman & Cooper, 1991c), mirror-reflection (Biederman & Cooper, 1991b), and surface characteristics (Biederman & Ju, 1988). In all cases they report no effects of a change in stimulus attributes as long as the same parts remain visible.

The role of outline shape in object recognition

In this paper, the sufficiency of outline shape information for recognition is investigated. Outline shape is defined simply as the bounding contour of an object from a particular viewpoint. The methodology employed uses both depth-rotations and changes in visible information from single viewpoints. Of interest is whether changes in outline shape predict recognition performance for both silhouettes and images containing more information. Such a finding would suggest that object recognition processes are based, at least partially, on outline shape information.

These results would not necessarily favour one of the models discussed above. Part-based, viewpoint-invariant models might, with an appropriate algorithm for extracting a structural description from the outline, predict similar performance for silhouettes and more realistic stimuli, as long as appropriate parts were visible in the silhouette. Similarly, viewpoint-dependent models might predict that normalization processes can match two views of an object on the basis of outline shape alone, and so might also predict little cost to recognition performance. Crucially, a finding that changes in outline shape predict costs to recognition would need to be addressed by any theory of object recognition.

Why might silhouettes provide useful information for recognition processes? There are at least two reasons. First, outline shape is presumably computed by figure-ground processes. Although the task of differentiating figure from ground is not a trivial one, and may in turn be influenced by object familiarity (Gibson & Peterson, 1994; Peterson & Gibson, 1993), it may be easier to compute than, say, the determination of whether an internal contour is shape-based (for example, the edge of one part which occludes portions of another part) or surface-based (for example, a pattern on the surface of an object). Second, large features of the object will tend to occur in the outline shape, and so processing silhouette information may be a rough way of determining crucial shape information from the object. This method will fail when a set of objects have the same regular shape (such as the discrimination of a television from a microwave oven), but these will likely form a small minority of object recognition decisions (for example, Tversky & Hemenway, 1984, argue that objects in different basic level categories tend to have different sets of parts, which will lead them to have different shapes).

Relatively few studies have investigated recognition of objects via outline shape previously. Research in the area of word recognition shows that in at least some situations, words are recognized on the basis of their outline shape (Crowder, 1982; Haber & Haber, 1981; Taylor & Taylor, 1983; but see also Besner, 1989). Other research has examined the interpretation of silhouettes as three-dimensional figures (Richards, Koenderink, & Hoffman, 1987; Schiano, McBeath, & Bruner, 1994; Walker & Walker, 1988) and in making figure-ground judgments (Gibson & Peterson, 1994). Rock, Halper, and Clayton (1972) show that recognition of novel figures is based more on outline shape than on the internal contours of an image. None of these studies, however, directly compares human performance at recognizing silhouettes and realistic images of stimuli, or examines whether effects of viewpoint change are predicted by changes in outline shape.

As a starting point, it seemed important to examine recognition based on outline shape in a situation in which recognition based on the entire visual image had previously been documented. Experiment 1, therefore, is a partial replication of Biederman and Gerhardstein's (1993) Experiment 3. Subjects in Biederman and Gerhardstein's experiment saw stimuli that might be rotated from a previous viewpoint. If rotated, costs to recognition were observed only if the rotation resulted in a different set of visible parts. Of interest in the current context was whether similar results would occur when the second presentation of the stimulus showed a silhouette. If so, outline shape information must be sufficient for recognition.

Experiment 1

Experiment 1 used stimuli similar to those of Biederman and Gerhardstein (1993). In their Experiment 3, novel objects were created which were composed of five volumetric shapes. Subjects saw two images in quick succession, and judged whether they had seen two presentations of the same object or presentations of two different objects. Drawings of the stimuli were made from three different viewpoints. The viewpoints were created in such a way that a rotation from the middle viewpoint in one direction produced an image of the object that showed the same parts to the previous viewpoint, whereas a rotation of the object in the opposite direction produced an image that had different visible parts. This manipulation also affected the outline shape of the object; the same-parts viewpoint showed roughly a mirror-image of the outline shape, whereas the different-parts viewpoint produced a marked change in outline shape. Many studies (e.g., Biederman & Cooper, 1991b; Corballis, Zbrodoff, Shetzer, & Butler, 1984; Tarr & Pinker, 1989) have found costs to recognition performance following mirror-reflection of a stimulus are trivial. Thus, recognition of the mirror-image silhouette is predicted to show similar latencies and accuracy as recognition of the silhouette of the previously observed viewpoint.

Subjects in this experiment viewed shaded images and silhouettes of objects similar to those used in Biederman and Gerhardstein's experiment. Two images were presented in succession; the first was always a shaded rendering of an object, and the second was either a shaded rendering or a silhouette. Of interest was whether performance for the silhouettes was different from performance for the shaded images. If recognition processes used for the stimuli in this experiment required information about internal contours or surfaces, recognition performance for the shaded images should be quite different to recognition of silhouettes. If such processes can use outline shape information alone, performance should be similar across both types of stimuli.

Method

Subjects. Thirty undergraduates at Yale University participated in the experiment in return for course credit. The data from one subject was excluded from the analysis, as accuracy was at chance.

Figure 1 about here

Stimuli. Ten objects were used in the experiment. All were based on line drawing stimuli that were employed in Biederman and Gerhardstein's (1993) Experiment 3. Each object consisted of a vertically-oriented central volume, and four smaller horizontally-oriented volumes that came out of the central volume. The stimuli used in this experiment are shown in Figure 1a. Three-dimensional models of each object were created using Alias Sketch! (Alias Research Inc: Toronto) on a Macintosh computer. Three viewpoints (A, B, and C) of each object were used, examples of which are shown in Figure 1b. Viewpoints A and B were 45° apart, and were roughly mirror images of each other, reflected in the vertical axis. Viewpoint C was also a 45° rotation from viewpoint B but in the other direction from viewpoint A. As discussed above, viewpoints A and B retained both the same visible parts and a mirror image of the outline shape, whereas viewpoints B and C showed different parts and had a different outline shape.

Two different versions of each viewpoint of each object were used. The first was a shaded rendering of the object, which created an image, complete with shading, surface texture and contour information, but without cast shadows. The second version was created by setting all pixels that were not background in each shaded image to black. This process created silhouettes of each image. The images were all similar sizes, with the maximum height of 5.5 cm (5.25° of visual angle) and maximum width 7.6 cm (7.25° of visual angle).

All stimulus presentations were followed immediately by a mask. The mask was created by combining image features from different objects in the experimental set. Thus, it retained some of the characteristics of the experimental object set, but looked very different from the other objects.

Design. There were two independent variables in Experiment 1: Viewpoint Difference between the two presentations, and Image Type of the second stimulus. All trials contained one presentation of viewpoint B. The other presentation could be identical (B - B), a different viewpoint that kept both parts and outline shape intact (B - A or A - B), or a different viewpoint that changed both parts and outline shape (B - C or C - B). Across all trials, the first presentation of the pair was a shaded image, and the second presentation was either a shaded image or a silhouette. There were 10 objects. Half the trials were targets, in which both presentations were of the same object; in the other half, trials consisted of two different objects. Finally, there were two blocks of trials, as all trials were repeated once. This created 240 total trials. Different trial orders were randomly created for each block for each subject.

Procedure. Subjects were instructed that they would see one object appear briefly, followed by a mask, and then a second object, again followed by the mask. Subjects were informed that their task was to judge whether the two presentations were of the same object or different objects. They were informed that objects could be shown from different viewpoints, and that they should respond only on the basis of object identity. Subjects were also informed that one object would often be a silhouette. They received no practice trials.

The experiment used a sequential matching task. Each trial was initiated by the presentation of a fixation cross for 500 ms on the center of a Macintosh CRT. The first object was presented for 200 ms, followed by the mask for 750 ms. The second object was then presented for 100 ms, and then masked for 500 ms. Subjects were instructed to respond as soon as they saw the second object; thus responses were recorded from the onset of the second object. A response key was pressed only if subjects judged the second object to be the same as the first; if they judged it to be different, they were instructed to do nothing (a go-no-go task). Subjects were required to respond within 1500 ms.

Figure 2 about here

Results

Mean reaction times for correct responses and overall error rates are shown in Figure 2. Errors represent failures to respond within 1500 ms. Analyses of variance (ANOVAs) were performed with two factors, Image Type of the second exemplar (shaded image or silhouette) and Viewpoint Difference between the stimuli (same viewpoint as first exemplar, different viewpoints that showed the same parts, or different viewpoints that showed different parts). For recognition latencies, only the factor of Viewpoint Difference was reliable, $F(2,56)=20.75$, $p<.001$. The factor of Image Type, $F(1,28)=1.96$, n.s., and the interaction between Viewpoint Difference and Image Type, $F(2,56)=1.98$, n.s., were both non-reliable. These results were not due to any effects of differential performance across the objects, as items analyses, using objects as the random factor, showed the same results (Viewpoint Difference: $F(2,18)=34.84$, $p<.001$; Image Type: $F(1,9)=.09$, n.s.; Viewpoint Difference by Image Type: $F(2,18)=1.76$, n.s.). For both the shaded images and the silhouettes, identical viewpoints were recognized fastest, followed by viewpoints that retained the same parts and outline shape, and finally viewpoints that changed the visible parts and the outline shape. All differences for both picture types between these three groups were reliable on post-hoc Scheffé tests (post-hoc tests, here and elsewhere, were conducted on subjects), with the exception of the difference between the identical rotation and same part rotation conditions for the silhouettes. This difference was marginally significant, at $p<.054$. Here, then, unlike Biederman and Gerhardstein's Experiment 3, a rotation that retained the same parts impaired recognition. A viewpoint that occluded parts, however, resulted in a greater impairment of recognition than did a similar sized rotation in depth that showed the same visible parts from the initial presentation.

For errors, both main factors and the interaction were reliable (Viewpoint Difference: $F(2,56)=25.82$, $p<.001$; Image Type: $F(1,28)=4.62$, $p<.05$; Viewpoint Difference by Image Type: $F(2,56)=7.43$, $p<.005$). The general pattern of errors follows the latency data, in that objects presented in the same viewpoint as previously were most accurately recognized, followed by viewpoints that showed the same parts and outline shape, and finally viewpoints that changed the visible parts and outline shape. The main effect of Image Type was caused by less accurate judgments of the silhouettes in two of the three rotation conditions; identical viewpoint, and rotation that retained the same visible parts. The interaction was reliable because this reduction in accuracy for judgments of silhouettes did not occur over all three rotation conditions; there was no difference in error rates between shaded images and silhouettes for rotations that produced a change in the visible parts. For items analyses, only the main effect of Viewpoint Difference was reliable, $F(2,18)=10.21$, $p<.001$; the main effect of Image Type, $F(1,9)=1.35$, n.s., and the interaction of Viewpoint Difference and Image Type, $F(2,18)=1.34$, n.s., were both non-reliable, suggesting that the reduction in performance for the two silhouette conditions did not occur robustly across all objects.

Discussion

The most salient pattern in the latency and error results for both types of image is that recognition performance was best for identical images, followed by rotations that kept the same parts and outline shape visible, and was worst for rotations that changed the visible parts and outline shape. For both measures, recognition performance was reduced for the images that showed a different collection of visible parts to the prior presentation (whether the parts were completely visible, as in the shaded images, or were only visible in the silhouette).

Differences between two sets of means in these data are worth noting. First, for both latencies and errors, recognition of silhouettes was slightly inferior to recognition of shaded images (although this pattern was only statistically significant for the accuracy data with subjects as the random factor). This result suggests that silhouettes may provide slightly less information for recognition processes than fully shaded images. Second, costs to recognition, in terms of both recognition latencies and accuracy, were not predicted by changes in the visible part structure across a rotation. For the shaded images, recognition was both faster and more accurate when the second viewpoint was exactly the same as the first, compared to situations when the second viewpoint was a rotation that kept the same parts visible. For silhouettes, the same pattern was reliable for error rates, and marginally significant for latencies. It appears that changes in viewpoint can affect recognition even if the same parts are visible in both viewpoints.

The results of this experiment suggest that in a sequential matching task with a short, masked interstimulus interval, the outline shape of a stimulus provides enough information for accurate identification. Given the short interval over which information had to be retained, however, it seemed important to examine the effects of recognizing shaded images and silhouettes in a different task. Many previous experiments used to examine visual object recognition (Bartram, 1974; Biederman & Cooper, 1991a; Biederman & Gerhardstein, 1993; Srinivas, 1993) employ naming tasks with intervals of at least 10 minutes between study and prime tasks, and Srinivas (1995) has shown large differences in recognition performance across different tasks. If outline shape is a general property used in visual image processing, one would expect it to mediate name priming in a task with a considerable lag between presentations of an object. Experiment 2 was therefore conducted to examine differences in object priming between silhouettes and realistic images of stimuli, across a markedly longer interval than was used in Experiment 1. In order to use a naming response, familiar objects were used as stimuli.

Experiment 2

Method

Subjects. Forty Yale University undergraduates participated in Experiment 2 in return for course credit. None had participated in Experiment 1.

Figure 3 about here

Stimuli. Objects taken from forty entry-level categories (Jolicoeur, Gluck, & Kosslyn, 1985) were used in this experiment. For 20 categories, one type of object was used, while for the other 20 categories, two different exemplars were used. For each of these 60 objects, a three-dimensional model was manipulated using Alias Sketch! or Swivel 3D (VPL Research, Inc.). Shaded renderings of each object were created in a viewpoint that was judged canonical by the experimenter. Silhouettes were then created from each of these renderings by turning each pixel in the image that was not background to black. Thus 60 realistic images and 60 silhouettes were used in the experiment. The images were similar sizes to each other, although there was some variation; the maximum height was 12 cm (11.4° of visual angle) and the maximum width was 18 cm (17° of visual angle). Examples of the stimuli are shown in Figure 3.

As in Experiment 1, a mask was created by combining image features from different objects in the experimental set. Thus, it retained some of the characteristics of the experimental object set, but looked very different from the other objects.

Design. There were two independent variables: Object Type (same or different) and Image Type (shaded or silhouette) of the second trial of each pair. Half the categories included two objects. For these, subjects named one in the first set of trials, and the other in the second set. As subjects were providing the same name for different stimuli (e.g. “car” for two different cars), this manipulation provided a baseline of conceptual and phonological priming (Biederman & Cooper, 1991a). The other half of the categories had only one object. The objects in these categories were named in both sets of trials.

All images presented in the first set of trials were shaded images of objects. During the second set of trials, half the objects were shaded images, whereas the other half were silhouettes. Half of the objects shown in the second set were the same as the objects presented in the first set of trials. The other half were different exemplars with the same name as an object shown in the first set. All objects served as silhouettes for half the subjects, and shaded renderings for the other half. In both study and test blocks there were 40 trials, and the order of presentation in each block was varied randomly for each subject.

Procedure. The experiment consisted of two sets of trials. At the start of each set, subjects received three practice trials. On each set, subjects viewed a series of images, and vocally named the presented objects. Subjects were told to use the first name that came to mind for each object. On the first set, each trial was initiated by a fixation cross, which appeared for 500 ms. The object then appeared for 500 ms, followed by a mask that remained on the screen until the subject responded. Each response was spoken into a microphone. The microphone was connected to the computer through a CMU Button Box (Carnegie Mellon University). This enabled the computer to measure the latency of the response. Response latencies were taken from the beginning of presentation of the object to the moment at which the response was made. Each response initiated the next trial after an intertrial interval of about a second.

At the completion of the first set of trials, subjects played a computer game as a distractor task for ten minutes. The second set of trials was then presented. Subjects were told that some of the objects would be familiar, and that some would be silhouettes. They were also instructed to name all the objects as quickly as they could. The only difference between the two sets of trials was that the images were each presented for 200 ms in the second set.

Results

Naming latencies were calculated only for those objects that were named without hesitation and with enough intensity to activate the voice trigger, were given the same name on both the first and second presentations, and were given a name which represented an entry-level classification for the object (e.g. trials were rejected if a cow was named as "animal"). Further, responses were also rejected if the naming latency of an object in either block was longer than 2000 ms. This procedure resulted in the omission of 21.5% of trials. Upon closer examination of the individual objects used in the experiment, it was found that for eight objects (one from the "same" group, and seven from the "different" group), trials were omitted on at least 50% of cases. When these objects were excluded, the omission rate for the 52 remaining objects was 12.25%. In all further analyses, results described exclude these eight objects for which performance was poor. In all cases analyses showed the same results for the complete data set. There was a reliable difference in the rejection rates of trials in which the same object was shown twice and trials in which two different objects were presented (15.9% versus 8.6% respectively; $\underline{F}(1,39)=14.7$, $\underline{p}<.001$), although this difference was not reliable in the items analysis, $\underline{F}(1,50)=.31$, n.s., showing that, even after excluding objects for which performance was very poor, the omission rate was still somewhat variable across objects. There were no reliable differences in the exclusion rate between shaded images and silhouettes, $\underline{F}(1,39)=1.31$, n.s., nor was the interaction between Object Type and Image Type significant, $\underline{F}(1,39)=1.03$, n.s. Thus, slightly more trials were omitted from the same object condition than the different object condition (although this difference was only reliable across analyses by subject), but no more were omitted for silhouettes than shaded objects. Given that different objects were used in the two Object Type conditions (same or different), but all objects could appear as shaded images or silhouettes, these results suggest performance was more variable for those objects in the same object condition, but was no more variable for silhouettes than shaded images. As this rejection rate includes hesitations, failures to record the response, and inconsistencies in names used, it should not be considered an error rate.

Amount of priming was calculated for the remaining trials by subtracting, for each subject, the naming latency for an object on the second set of trials from the latency for the same object (or object with the same name) on the first set. This procedure produced the extent that each subject was faster in naming each object on the second block relative to the first block. The average naming latency on the first set of trials was 951 ms; the average latency on the second set was 900 ms.

Figure 4 about here

Priming is shown in Figure 4. Substantial priming was found for objects that were preceded by themselves, but much less priming was found for objects preceded by a different object with the same name. Differences in naming latency between the first and second block were only reliable when the object primed itself (shaded images: $\underline{t}(308)=8.86$, $\underline{p}<.001$; silhouettes: $\underline{t}(294)=5.59$, $\underline{p}<.001$), and were not reliable when the object primed a different exemplar of the same category (shaded images: $\underline{t}(309)=0.95$, n.s.; silhouettes: $\underline{t}(256)=1.52$, n.s.). Typically, priming studies report reliable priming for both identical and different exemplars within the same category (e.g. Bartram, 1974; Biederman & Cooper, 1991a; Biederman & Gerhardstein, 1993; Warren & Morton, 1982). It is unclear why there was no priming effect in this experiment for different objects with the same name as a previously presented stimulus. Crucially, however, there are large differences in performance between identical and different exemplar conditions, suggesting that subjects show purely visual, non-semantic, non-phonological priming for identical stimuli (Biederman & Cooper, 1991a).

There was little difference in the amount of priming for shaded images and silhouettes. These observations are supported by an ANOVA, which was performed with amount of priming as the dependent variable and two independent variables, Object Type (same or different) and Image Type (shaded image or silhouette). Only Object Type was reliable, $\underline{F}(1,39)=22.64$, $\underline{p}<.001$. Neither the main effect of Image Type, $\underline{F}(1,39)=1.6$, n.s., nor the interaction between Object Type and Image Type, $\underline{F}(1,39)=2.3$, n.s., were reliable. The effect of Object Type was marginally reliable on an items analysis, $\underline{F}(1,50)=3.46$, $\underline{p}<.07$. Visual object priming was found, as an object was primed more by the prior presentation of itself than by a different object that had the same name. For just those trials in which the second presentation showed the same object as the first presentation, silhouettes showed 37 ms less priming than the shaded images, but this difference was not significant on a post-hoc Scheffé test ($\underline{p}>.09$).

Discussion

Using different objects and a different task, the results of this experiment are very similar to those of Experiment 1. Facilitation was found in naming performance whenever the second object was the same as the first, regardless of whether it was presented as the same shaded image or a silhouette. There was a small difference in the amount of facilitation between the two image types, but this difference was only marginally reliable. Thus, performance for fully shaded images may be slightly better than performance for silhouettes, but this small difference was not similar to the dramatic cost of showing a second shaded image that was a different exemplar of the object. An account of object recognition that used only information available in the outline shape of a stimulus would predict all the statistically reliable results that were found in this experiment.

One drawback with the design of the experiment was the inclusion of different objects in the two Object Type conditions (i.e. objects were only ever in the same exemplar or different exemplar conditions). This design was necessary due to the small number of categories for which two different objects were available. Because of this design, it is possible that some differences in priming between the same and different exemplar conditions are due simply to the stimuli. Crucially, however, differences between the two sets of objects are unlikely to affect the interpretation of the main effects in this experiment. The main effect for Object Type replicates a consistent body of evidence (e.g. Bartram, 1974; Biederman & Cooper, 1991a; Biederman & Gerhardstein, 1993; Warren & Morton, 1982) that finds specifically visual priming from one object to another. The null effect for Image Type could not be due to differences in stimuli between the same and different exemplar conditions.

Summary of Experiments 1 and 2

Experiments 1 and 2 find little difference in the recognition of silhouettes compared to the recognition of shaded images of objects. In Experiment 1, silhouettes were matched as fast and almost as accurately as shaded images. Silhouettes also showed the same pattern of performance across changes in viewpoint as shaded images. Experiment 2 found no statistical difference in the amount of visual priming for shaded images and silhouettes.

There are two constraints on the interpretation of results from these two experiments. First, the objects used in both studies were of widely varying outline shapes, and so this fact may have encouraged subjects to rely on information present in the silhouette when performing the recognition and naming tasks. Although this concern needs to be acknowledged, the objects were similar to those used in a wide range of previous experiments. If a concern about similarity of outline shape constrains the interpretations of these results, it should also constrain the interpretation of other experiments. Second, these results do not imply that other information about objects was not encoded. In both experiments, there were small differences in recognition of shaded images and silhouettes, and these differences may be due to processing of non-outline shape information for the shaded images.

Despite these concerns, two conclusions can be drawn from the results of these experiments. First, subjects can perform these tasks on the basis of outline shape alone. Second, changes in outline shape can be used to accurately predict performance in these tasks, especially in Experiment 1 when the object is rotated in depth. It appears that any account of performance on both naming tasks and sequential matching tasks must be able to explain object recognition on the basis of the outline shape of the stimuli.

Experiment 3

Both of the experiments that have been reported used silhouettes to test for the role of outline shape in recognition processes. One problem with this approach is that by seeing silhouettes, subjects may have been primed to use outline shape to perform the required tasks. Given that half of the objects in both test situations were silhouettes, this use of information would have made strategic sense. Such a possibility undermines the generality with which outline shape will be used in everyday object recognition. In order to establish that the use of outline shape in Experiments 1 and 2 was not simply due to the use of silhouettes, a different set of images was used. In Experiment 3, all stimuli were realistically shaded pictures of objects.

The stimuli were created by rotating objects in depth. As shown in Figure 5b, a rotation of 180° will show an outline of the object that is largely a mirror image of the outline shape at the original viewpoint (when objects, as here, are shown in perspective projection). Such a rotation disrupts much visible information and might be predicted by theories of object representation that rely on parts (e.g. Hummel & Biederman, 1992) or orientation difference between viewpoints (e.g. Tarr & Pinker, 1989) to impair performance on recognition tests. If recognition can be based on the outline shape of an object, however, performance on a recognition task might be better at a 180° rotation than at a smaller rotation in which the outline shape of an object is less similar to its appearance at the original viewpoint. Experiment 3 therefore used rotations that would vary the similarity of outline shape between two viewpoints, while using a similar task to that of Experiment 1. As this experiment did not use silhouettes, it is less likely than the previous experiments to make the outline shape of the objects unusually salient. Experiment 3 used a similar sequential matching task to that employed in Experiment 1.

Method

Subjects. Twenty-one Yale University undergraduates participated in the experiment in return for class credit. None had participated in any of the earlier experiments.

Figure 5 about here

Stimuli. Twenty familiar common objects were used in Experiment 3, and are shown in Figure 5a. Each object was realistically rendered at three viewpoints, examples of which are shown in Figure 5b. The first was produced by taking the viewpoint in which the axis of elongation was running from the viewer to the horizon, and the intrinsic front of the object (if it had one) was facing the viewer, and then performing a 30° rotation. This viewpoint, as the baseline from which the others were rotated, was labelled 0°. The other two viewpoints were created by first, a 60° rotation from the starting viewpoint (with the rotation occurring away from the front, so as not to rotate the object across the front), and second, a 180° rotation from the starting viewpoint. All stimuli were grayscale images, so that subjects could not perform the match based on color. The images were of similar sizes to those used in Experiment 2.

The view of the objects was such that the 0° and 180° viewpoints showed roughly a mirror-image of outline shape. It also seemed important to define which viewpoints share visible parts, but this was much more difficult, because part-based models of object recognition (e.g., Biederman, 1987; Hummel & Biederman, 1992) do not necessarily specify which particular parts of an object will be crucial for recognition. The default assumption was therefore that all viewpoints show all crucial parts of each object. This assumption suggests that recognition should be invariant across viewpoints. It should be noted that some parts are occluded across rotations (such as, in Figure 5b, the front lights, bonnet, and windscreen of the car). However, because the 60° rotation showed a side of the object that was visible at 0°, these two viewpoints were assumed to show similar parts. On the other hand, the 180° viewpoint should be maximal at occluding information from 0°. Thus, it was assumed that if parts were occluded, recognition should be better at the 60° viewpoint (given the prior presentation of a 0° viewpoint) than at the 180° viewpoint.

The mask in this experiment was a series of irregularly shaped lines. Each line was a different shade of gray.

Design. There was only one independent variable in Experiment 3: Viewpoint Difference between the first and second presentations of the object. This difference could be 0°, 60°, or 180°. One presentation of each object was always from the 0° viewpoint, and the other was one of the three possible viewpoints. Half the trials consisted of distractors, in which different objects from the stimulus set were shown in each presentation. Finally, two random orders of presentation were created for each subject, so that each object was seen twice in each pairing. This created 240 trials per subject.

Procedure. The experiment was almost identical to Experiment 1, except that the stimuli were all realistically shaded images, and were all natural objects. Subjects were instructed that two images would be presented, and that they should judge whether the two presentations depicted the same object or two different objects. They were warned that objects could be shown from different viewpoints, and that they should respond only on the basis of object identity. They received no practice trials. Unlike Experiment 1, subjects responded with one key when the same object was presented twice, and another key when two different objects were presented.

The timings with which the different stimuli and masks were presented were the same as Experiment 1.

Results

Figure 6 about here

The recognition latency and error rates for trials on which the same object was shown at each presentation are shown in Figure 6[1]. Trials on which subjects took longer than 1500 ms to respond (9.4% of total trials) were omitted from the analysis. There is a clear ordering of performance, which is consistent in both sets of data. The objects shown in the same viewpoint in both presentations were recognized most accurately and fastest. The objects which underwent a 60° rotation between the two presentations were recognized least accurately and slowest. Performance across a 180° rotation fell between these extremes. The pattern of omissions (responses not made during the 1500 ms window) followed the RTs and error rates; 8.6% of responses in the 0° condition were omitted, 10.5% were omitted from the 60° rotation, and 9.1% were omitted from the 180° rotation.

These conclusions are supported by ANOVAs that were run on the latency and error data. In each, there was a reliable main effect for Viewpoint Difference (Recognition Latency: $F(2,40)=16.12$, $p<.001$; Errors: $F(2,40)=13.57$,

---

[1] By way of comparison, mean RT for the distractors was 695 ms, and the mean error rate was 28%.

$p<.001$). These analyses were also reliable across items (Recognition Latency: $F(2,19)=11.84$, $p<.001$; Errors: $F(2,19)=9.83$, $p<.001$). Post-hoc Scheffé tests show reliable differences at the .05 significance level between all three conditions in the recognition latency data, and reliable differences in the error data between the 0° and 60° groups, and the 180° and 60° groups.

Discussion

Recognition performance was best when the object was identical in both presentations, a finding which mirrors Experiment 1. When the object was rotated in depth, however, a rotation that retained a similar outline shape was recognized faster and more accurately than a smaller rotation that changed the outline shape. The 180° rotation did not appear to share more visible parts with the initial viewpoint than the 60° rotation shared with the initial viewpoint, and if anything appeared to share less. The differences found in this experiment are therefore difficult to explain with reference to a parts-based structural description. This type of theory, exemplified by RBC, predicts inferior recognition performance when a second viewpoint shows different parts to the first viewpoint. At the very least, performance might be expected to be equivalent in the 60° and 180° rotations, if they both shared the same visible parts as the 0° viewpoint. One possible explanation, to be addressed in the General Discussion, is that performance might be mediated by parts that are visible in the outline shape of an object from a particular viewpoint.

The results of Experiment 3 also show that this task is not performed by a mechanism in which recognition costs are based on a linear function of the degree of rotation between the viewpoints, or the 180° rotation would be recognized slower and less accurately than the 60° rotation. Recent discussions of viewpoint-dependent models by Tarr and his colleagues (Hayward & Tarr, in press; Tarr et al., 1994; Tarr & Bülthoff, 1995; Tarr & Kriegman, 1993) have proposed that representations might be normalized on the basis of viewpoint-specific image features, in particular (although not exclusively) those shape features of the outline shape of an object. In addition, Vetter et al. (1994) have shown that subjects may represent previously unseen, “virtual” views on the basis of available geometric information. Thus, viewpoint-dependent models might, in some circumstances, predict non-linear normalization from previously presented viewpoints. Identifying the specific image features that are normalized across changes in viewpoint is beyond the scope of the current paper, however.

There is one possible concern with the design of Experiment 3. All objects were viewed at the 0° viewpoint, along with one other (0°, 60°, or 180°). Thus, in this experiment the size of rotation between viewpoints (i.e. whether the object was rotated between presentations by 0°, 60°, or 180°) is confounded with the viewpoint from which the object, on one presentation, was observed (i.e. whether it was observed from 0°, 60°, or 180°). The argument made in this paper is that faster and more accurate recognition for the 180° rotation, as compared to the 60° rotation, shows the effects of greater similarity in outline shape between 0° and 180° than between 0° and 60°. This argument rests on the assumption that from the 180° viewpoint the object is no easier to recognize than it is from the 60° viewpoint. In order to ensure that this assumption was indeed true, a control experiment was run. Fifteen subjects were presented with a similar sequential matching task to that used in Experiments 1 and 3, except that the first stimulus was always a category name, and the second stimulus was an object from Experiment 3 shown from either 0°, 60°, or 180°. If the objects were more easily recognized from 180° than from 60°, recognition in this experiment should be faster at 180°. In fact, subjects were equally fast to categorise the objects from 0° and 60° (604.9 ms and 604.8 ms respectively), and were slower to recognize the objects from 180° (623.1 ms), $F(2,28)=3.36$, $p<.05$. Despite the difficulty of recognizing the objects from 180°, therefore, in Experiment 3 subjects were more easily able to match this viewpoint than a 60° viewpoint to an initial presentation.

Experiment 4

As with Experiment 1, the task employed in Experiment 3 required only very short retention of the first image (750 ms). Although the use of outline shape was found in this experiment, such use may still be an artefact of the short term nature of the task. Experiment 4 therefore uses the longer term, naming paradigm from Experiment 2 to explore object priming, but uses the same type of stimulus rotation as Experiment 3. This manipulation is again designed to address the concern that previously, subjects used outline shape simply because it was salient. In Experiment 4, subjects first named a series of objects presented at the previously defined 0° viewpoint, and then later named a set of objects rotated 60° and 180° from this initial study viewpoint.

Method

Subjects. Forty Yale University undergraduates participated in Experiment 4 in return for class credit. None had participated in any of the previous experiments.

Stimuli. The stimuli were 32 realistically shaded natural objects, similar to those of Experiment 3. Each object was rendered at the three rotation conditions used in Experiment 3; 0° (which was roughly 30° from front), 60°, and 180°. The front of the object was defined in the same way as for Experiment 3. All objects used in this experiment were from different entry level categories. Stimuli were presented in color. The images were of similar sizes to those used in Experiment 2.

The mask in this experiment was the same as that used in Experiment 2.

Design. There were two independent variables; Viewpoint Difference and Study Condition. All objects were presented at 0° during the first set of trials, and at either 60° or 180° during the second set. In the second set of trials, objects were either previously studied or unstudied. For each subject, half the objects were presented in both sets of trials, and half were presented only in the second set of trials. This manipulation ensured a baseline from which to measure priming. All objects served in all stimulus conditions across different subjects. The order with which stimuli were presented was randomly varied for each subject.

Procedure. The procedure was similar to Experiment 2. There were two main differences. First, all stimuli were realistically shaded images. Second, a different type of baseline was calculated. Of interest in this experiment was the degree to which subjects would be faster in naming an object at a particular viewpoint when they had previously seen the object at 0° relative to naming an object at that viewpoint when it had not been previously studied. The appropriate baseline was therefore the time taken to name previously unstudied objects from the target viewpoints (i.e. 60° and 180°). Thus, unlike Experiment 2, there were no objects presented which were different exemplars but from the same entry level category as earlier stimuli.

Subjects were initially instructed that they would see objects appear in succession, and that they should name each one into the microphone. They received three practice trials before each set. Sixteen objects were named in the first set of trials, and then subjects played a computer game for 10 minutes as a distraction. Subjects then named all 32 pictures in the second set of trials. The timings of all stimuli in each trial were identical to those used in Experiment 2.

Results

As in Experiment 2, recognition latencies were calculated only for those objects that were given the same name on each trial without hesitation, when those names represented an entry-level classification for the object, and when responses on the second block were less than 2000 ms. In similar fashion to Experiment 2, six objects had omission rates of at least 50%. All reported analyses omit data from these six objects. All statistically significant results were also reliable across the total data. Of the data included in the reported analyses, 13% of trials were omitted for one of the previously noted reasons. There were no reliable differences for the omission rate across the different conditions of the experiment.

Figure 7 about here

There are differences in the ease with which objects are named from different viewpoints, (e.g., Palmer, Rosch, & Chase, 1981, and the control condition from Experiment 3). Thus an initial hypothesis was that objects in the 180° viewpoint might be generally more difficult to name, as they primarily showed the back of the object. Priming, in this experiment, was therefore the time taken to name an object that had been previously viewed at a particular viewpoint, relative to the latency of naming in that same viewpoint an object that had not been previously viewed. Naming latencies are shown in Figure 7. Supporting the initial hypothesis, there is a large difference in the naming latencies for non-studied objects at 60° and 180°, suggesting that 180° viewpoints of an object are relatively difficult to name. The 60° condition shows relatively little facilitation from the prior presentation of an object at 0°. Objects in the 180° condition, on the other hand, were named considerably faster if the object had been previously viewed at 0° than if it had not been previously viewed.

An ANOVA confirms these observations. Both main effects of Study Condition, $F(1,39)= 17.9$, $p<.001$, and Viewpoint Difference, $F(1,39)=21.6$, $p<.001$, were reliable. It appears, however, that the main effects are due to the relatively large difference in recognition latencies for the 180° rotation, as shown by a reliable interaction between these factors, $F(1,39)=8.5$, $p<.01$. These effects were also reliable on items analyses (Study Condition: $F(1,25)=19.2$, $p<.001$; Viewpoint Difference: $F(1,25)=11.0$, $p<.01$; Study Condition by Viewpoint Difference: $F(1,25)=12.9$, $p<.01$).

Discussion

Facilitation for objects shown at 180° by the prior presentation of the same object at 0° was much larger than any similar facilitation for the objects at 60°. As discussed above, objects shown at 180° did not appear to share more parts with the view of the object at 0° than the 60° viewpoint shared with 0°. The outline shape of the objects at 0° and 180° was relatively similar, however. The larger degree of facilitation in this latter case suggests that the outline shape of the previous viewpoint of each object was used to aid the naming of the object in the second set of trials.

One difficulty in the interpretation of the results of Experiment 4 is the possibility of a floor effect. If subjects were unable to name the stimuli any faster than about 900 ms, the observed pattern of results would be found, but one would not be able to conclude that the pattern was due to processing differences. Two findings suggest that these results are not due to a floor effect. First, naming latencies of 900 ms are relatively slow (compared to naming latencies in other studies; see for example Biederman & Cooper, 1991a; Biederman & Gerhardstein, 1993; Jolicoeur, 1985), and so performance is unlikely to be optimal in any of the experimental conditions. Second, a control condition was run to investigate whether any differences could be found between stimuli that had been previously presented. If the results of Experiment 4 are due to a floor effect, any prior presentation of an object should push naming latencies to floor. Twenty subjects named the objects that had been used in Experiment 4 at the 60° viewpoint, and then later named the same objects, half of which were presented at 60°, with the other half presented at the 180° viewpoint. If performance on this task is pushed to floor following any prior presentation of the same stimulus, naming latencies should be identical for these two groups of stimuli. Naming was reliably faster for the identical viewpoint (834 ms) than for the rotated viewpoint (888 ms), $\underline{F}(1,29)=10.97$, $\underline{p}<.01$, showing that facilitation differences were possible for this task. It appears unlikely, therefore, that performance on the facilitated stimuli in Experiment 4 was due to a floor effect.

General Discussion

In each of the four experiments, similarities in outline shape of the presented stimuli predicted recognition performance. Experiments 1 and 2 contrasted recognition of shaded images with recognition of silhouettes, and found that performance for silhouettes was very similar to, and in most cases indistinguishable from, recognition performance for shaded images. In Experiment 3, the recognition of a stimulus was easier if it shared outline shape with an earlier presentation, and more difficult if the outline shape was dissimilar. In Experiment 4, a similar manipulation affected facilitation on a naming task; more facilitation was gained if the outline shape was similar to a previous presentation than if the outline shape was different.

Implications for theories of object recognition

An open question is whether subjects in these experiments represented non-outline shape information, such as surfaces and internal contours, for the objects they studied. Performance in Experiments 1-3 was best for conditions on which the same stimulus was observed twice, suggesting that performance was aided when other, non-outline information was available. However, similarities in outline shape appeared to predict performance whenever the object was presented from a new viewpoint. Thus, outline shape might be particularly crucial for achieving object constancy, or recognition across a change in observed viewpoint. Supporting this assertion, Hayward and Tarr (in press) found recognition performance for single-part objects rotated in depth was predicted by qualitative changes in the outline shape of an object. One way of achieving object constancy might be to normalize the outline shape of an object from a particular viewpoint so that it matches the outline shape of a previously presented viewpoint.

The results of the experiments presented here show that any theory of object recognition must allow for recognition to occur on the basis of outline shape. At the beginning of this paper, part-based approaches, characterized by Biederman's Recognition-By-Components theory (Biederman, 1987; Hummel & Biederman, 1992), were contrasted with the viewpoint-dependent approaches of, for example, Tarr (1995) and Bülthoff & Edelman (1992). Neither type of theory explicitly discusses the role of outline shape in recognition. The remainder of this paper will therefore examine whether either is compatible with the findings presented here.

Can Geon-Structural-Descriptions account for outline-based recognition?

Biederman's RBC theory proposes that object representations consist of geometric primitives which are derived from an object's parts, and the spatial relations with which those primitives are connected. Presumably, structural description models such as RBC (and its later variants; Hummel & Biederman's, 1992, JIM and Hummel & Stankiewicz's, 1996, hybrid model of object recognition) could account for outline-based recognition if structural primitives are able to be recovered from information in the silhouette of an object from a particular viewpoint. Biederman (1987) suggested that geons could be recovered from an image on the basis of five non-accidental properties of edges in the scene; collinearity, curvilinearity, symmetry, parallelness, and vertex terminations. He also noted that not all edges defining a geon would have to be present for recovery of that geon to occur; a subset of contours, as long as they specified the appropriate non-accidental properties, would suffice. Any object part that occurs in the bounding contour will likely show some, but not all, of the bounding contour for that individual part. Of interest, then, is whether geon recovery can proceed on the basis of silhouette information.

Certainly, the extent of a part that appears in the outline shape of an object will often provide enough information for the determination of collinearity, curvilinearity, symmetry, and parallelness. In addition, vertices will be apparent for terminations of two curves at a single point (an “L” vertex), although not for three or more (e.g. a fork, or an arrow). Of course, in most cases a single part will be specified by a small subset of its actual contour. In addition, there will be many “L” vertices which specify not the termination of two curves (typically, convex vertices), but the occlusion of one curve by another (for example, when one part runs in front of another; these points will tend to be concave cusps). Hummel (1994b), however, provides an algorithm which can differentiate convex vertices from concave cusps, showing that this concern may not be insurmountable.

Regardless of whether those properties specified by Biederman (1987) as useful to geon recovery are present in the outline shape of an object, there are some fundamental concerns about whether it is possible to recover geons on the basis of this information. While considerable evidence has been generated on the role of parts in object recognition (e.g. Biederman & Cooper, 1991a; Biederman & Gerhardstein, 1993), relatively little behavioral evidence has been obtained on the issue of geon extraction from visual stimuli. Most relevant research on this topic has been performed by Hummel and his colleagues using multi-layered connectionist networks. Hummel and Biederman’s (1992) JIM model provides some explication of the principles introduced by Biederman (1987). However, this model uses only a subset of the original non-accidental properties assumed to be crucial for geon derivation (vertices and 2-D axes of symmetry). This development does not seem to make JIM any more likely than RBC to be able to derive geons from outline shape; as noted above, only “L” vertices will be recoverable from the outline shape, and Hummel and Biederman themselves note that there is currently no algorithm to compute 2-D axes of symmetry. Hummel and Stankiewicz (1996) produce a more convincing demonstration of the parsing of a visual image into parts, but it is still unclear whether the non-accidental properties specified by Biederman (1987) are sufficiently powerful to enable recovery of geon identities from this parsed image.

Kurbat (1994) notes that for a number of reasons, including those discussed above, RBC and JIM are unlikely to be easily revised to produce robust recovery of geon structural descriptions from objects, although he does not specifically consider the issue of recovery from outline shape. In particular, he notes that using vertices for recognition will often lead to erroneous conclusions about 3-D structure and that many parts do not have a regular axis or a regular shape around an axis. Using only outline shape will likely make these problems more difficult to resolve, although not necessarily in every case. For example, the vertices apparent in the silhouette of an object will all be shape-based, rather than surface-based, and thus may produce more vertices that are useful to recognition than will processes which operate on a fully shaded image. However, recovering geons from the outline shape would not resolve the problem of the regularity of part axes and shapes. Finally, it will also be difficult to recover the spatial relations with which geons are connected from the outline shape, as relations such as “front-back” will be impossible to determine. It seems, therefore, that current instantiations of the geon structural description approach are difficult to integrate with the current findings of object recognition on the basis of outline shape.

Can multiple-view models account for outline-based recognition?

Typically, models of object recognition that propose viewpoint-specific representations and recognition via normalization (e.g. Bülthoff & Edelman, 1992; Tarr, 1995), propose that such normalization processes use input from the entire visual image. For example, Bülthoff and Edelman (1992; based on Poggio & Edelman, 1990) propose that unfamiliar viewpoints of an object are matched to familiar viewpoints through the linear combination of viewpoint-specific features in the observed image. One crucial issue, as yet unresolved, is an explication of the specific features that will be normalized in an image. For example, are some shape features easier to normalize than others? Equally, will color and shadowing on an object be normalized between views or will only shape attributes be normalized? Until such issues are resolved, it is difficult to assess whether these models, as currently formulated, could account for recognition using only outline shape information.

Unlike geon structural description approaches, however, multiple view models appear to be compatible with the results of the experiments reported in this paper. Such models predict that recognition performance will deteriorate as a function of stimulus dissimilarity. The experiments in this paper show that dissimilarity in outline shape, in particular, accounts for costs to recognition. Thus, normalization processes may operate specifically on shape information in the outline of an object.

This proposal does not by itself explain how the outline shape of an object is normalized between disparate viewpoints. However, recent work provides a framework within which to examine this issue. Tarr and colleagues (Hayward & Tarr, 1996; Tarr, 1995; Tarr et al., in press; Tarr & Kriegman, 1993) have proposed that viewpoint-dependent representations might involve the qualitative coding of object shape. This proposal is based on the work of Koenderink and Van Doorn (1979), who argued that qualitative similarities in the geometry of vertices might be represented by a single “aspect graph” (see Van Effelterre, 1994, for a review of such approaches). Hayward and Tarr (1996) have shown that with rotations in depth, qualitative changes in the outline shape of a single part object predict costs to recognition performance. Thus, normalization may occur on the basis of qualitative codings of outline shape.

Conclusion

Clearly, viewpoint-dependent theories must be made more explicit about both the specific information encoded in representations and the precise mechanisms by which different representations can be normalized before strong claims can be made as to the situations in which they do or do not accurately predict recognition. The shortcomings of RBC (Biederman, 1987) in terms of recovering a geon structural description, detailed by Kurbat (1994), and in terms of predicting costs to recognition performance (Hayward & Tarr, 1996; Tarr, et al., 1994, in press), emerge because RBC has transparent assumptions and clear predictions. Despite these concerns, however, multiple-view theories appear more likely than part-based, structural description approaches to explain a growing body of data, reported in this paper as well as elsewhere (e.g., Bülthoff & Edelman, 1992; Hayward & Tarr, 1996; Lawson & Humphreys, 1996; Srinivas, 1993; Tarr, 1995; Tarr, et al., 1994, in press), which show systematic costs to recognition with changes in viewpoint.

In sum, then, outline shape appears sufficient for recognition in the experiments of this paper. Models of object recognition must therefore be able to show robust recognition from solely outline information. In particular, recognition of depth-rotated objects in Experiments 1, 3, and 4 seems well predicted by changes in outline shape. These results suggest that normalization across viewpoints may be based on shape characteristics in the outline of an image.

References

Bartram, D. J. (1974). The role of visual and semantic codes in object naming. Cognitive Psychology, 6, 325-356.

Besner, D. (1989). On the role of outline shape and word-specific visual pattern in the identification of function words: NONE. The Quarterly Journal of Experimental Psychology, 41A, 91-105.

Biederman, I. (1987). Recognition-By-Components: A theory of human image understanding. Psychological Review, 94, 115-147.

Biederman, I., & Cooper, E. E. (1991a). Priming contour-deleted images: Evidence for intermediate representations in visual object priming. Cognitive Psychology, 23, 393-419.

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

Biederman, I., & Cooper, E. E. (1991c). Object recognition and laterality: Null effects. Neuropsychologia, 29, 685-694.

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

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for 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. Journal of Experimental Psychology: Human Perception and Performance, 21, 1506-1514.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. Cognitive Psychology, 20, 38-64.

Binford, T. O. (1971). Visual perception by computer. Paper presented at the IEEE Conference on Systems and Control, December 1971, Miami.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a 2-D view interpolation theory of object recognition. Proceedings of the National Academy of Sciences, 89, 60-64.

Cooper, E. E., & Biederman, I. (1993). Geon differences during recognition are more salient than metric differences. Paper presented at the 34th Annual Meeting of the Psychonomic Society, Washington, DC.

Cooper, E. E., Biederman, I., and Hummel, J. E. (1992). Metric invariance in object recognition: A review and further evidence. Canadian Journal of Psychology, 46, 191-214.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), Visual Information Processing (pp. 75-176). New York: Academic Press.

Corballis, M. C., Zbrodoff, N. J., Shetzer, L. I., & Butler, P. B. (1984). Decisions about identity and orientation of rotated letters and digits. Memory and Cognition, 6, 98-107.

Crowder, R. G. (1982). The Psychology of Reading. New York and Oxford: Oxford University Press.

Cutzu, F., & Edelman, S. (1994). Canonical views in object representation and recognition. Visual Research, 34, 3037-3056.

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.

Fiser, J., & Biederman, I. (1995). Size invariance in visual object priming of gray-scale images. Perception, 24, 741-748.

Gibson, B. S., & Peterson, M. A. (1994). Does orientation-independent object recognition precede orientation-dependent recognition? Evidence from a cuing paradigm. Journal of Experimental Psychology: Human Perception and Performance, 20, 299-316.

Haber, R. N., & Haber, L. R. (1981). Visual components of the reading process. Visible Language, *15*, 147-182.

Hayward, W. G., & Tarr, M. J. (in press). Testing conditions for viewpoint invariance in object recognition. Journal of Experimental Psychology: Human Perception and Performance.

Hummel, J. E. (1994a). Reference frames and relations in computational models of object recognition. Current Directions in Psychological Science, 3, 111-116.

Hummel, J. E. (1994b). Segmenting images at matched concavities with synchrony for binding. Technical Report 94-1, Shape Perception Laboratory, University of California, Los Angeles.

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

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

Jolicoeur, P. (1985). The time to name disoriented natural objects. Memory and Cognition, 13, 289-303.

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Picture and names: Making the connection. Cognitive Psychology, 16, 243-275.

Koenderink, J. J., & Van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. Biological Cybernetics, 24, 51-59.

Kurbat, M. A. (1994). Structural description theories: Is RBC/JIM a general-purpose theory of human entry-level object recognition? Perception, 23, 1339-1368.

Lawson, R., & Humphreys, G. W. (1996). View-specificity in object processing: Evidence from picture matching. Journal of Experimental Psychology: Human Perception and Performance, 22, 395-416.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of three dimensional shapes. Proceedings of the Royal Society of London, Series B, 200, 269-294.

Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), Attention and Performance IX. Lawrence Erlbaum: Hillsdale, NJ.

Peterson, M. A., & Gibson, B. S. (1993). Shape recognition inputs to figure-ground organization of three-dimensional displays. Cognitive Psychology, 25, 383-429.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize 3D objects. Nature, 343, 263-266.

Richards, W., Koenderink, J. J., & Hoffman, D. D. (1987). Inferring 3D shapes from 2D silhouettes. Journal of the Optical Society of America, A, 4, 1168-1175.

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

Rock, I., Halper, F., & Clayton, T. (1972). The perception and recognition of complex figures. Cognitive Psychology, 3, 655-673.

Schiano, D. J., McBeath, M. K., & Bruner, K. W. (1994). Orientation-constraining heuristics for interpreting figures as 3-D objects. Paper presented at the 35th Annual Meeting of the Psychonomic Society, St Louis, MO.

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

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

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

Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. Psychonomic Bulletin and 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? Journal of Experimental Psychology: Human Perception and Performance, 21, 1494-1505.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., & Blanz, V. (in press). To what extent to unique parts influence recognition across changes in viewpoint? Psychological Science.

Tarr, M. J., Hayward, W. G., Gauthier, I., & Williams, P. (1994). Geon recognition is viewpoint-dependent. Paper presented at the 35th Annual Meeting of the Psychonomic Society, St Louis, MO.

Tarr, M. J., & Kriegman, D. J. (1993). Toward understanding human object recognition: Aspect graphs and view-based representations. Unpublished manuscript, Yale University, New Haven, CT.

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

Taylor, I., & Taylor, M. M. (1983). The Psychology of Reading. London and New York: Academic Press.

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

Van Effelterre, T. (1994). Aspect graphs for visual recognition of three-dimensional objects. Perception, 23, 563-582.

Vetter, T., Poggio, T., and Bülthoff, H. H. (1994). The importance of symmetryand virtual views in three-dimensional object recognition. Current Biology, 4, 18-23.

Walker, J. T., & Walker, M. J. (1988). The apparent size of three-dimensional objects and their silhouettes: a solid-superiority effect. Perception, 17, 23-30.

Warren & Morton (1982). The effects of priming on picture recognition. British Journal of Psychology, 73, 117-129.

Footnotes

1. By way of comparison, mean RT for the distractors was 695 ms, and the mean error rate was 28%.

Author Note

William G. Hayward, Department of Psychology, University of Wollongong.

This research was conducted in partial fulfillment of the requirements for the PhD in the Department of Psychology at Yale University. Support was provided a Yale University Graduate Fellowship, and at the University of Wollongong by an Australian Research Council Small Grant. Additional support was provided by the Office of Naval Research contract number N00014-93-1-0305 to Michael J. Tarr. I would like to thank Michael Tarr for advice and guidance throughout the research project. I would also like to thank all those who provided constructive comments on earlier versions of this manuscript, particularly David Kriegman, Robert Crowder, Elizabeth Phelps, John Kihlstrom, Joseph Lappin, Rebecca Lawson, Pierre Jolicoeur, and an anonymous reviewer. Thanks also to Scott Yu for advice on computer graphics issues, Pepper Williams and Isabel Gauthier for constructive criticism, and James Servidea and James Rosoff for helping to collect the data.

Correspondence concerning this article should be addressed to William Hayward, Department of Psychology, University of Wollongong, Wollongong NSW 2252, Australia. Electronic mail can be sent to w.hayward@uow.edu.au.

Figure Captions

Figure 1. (a) The ten objects used in Experiment 1; (b) three viewpoints of one object, showing both shaded image and silhouette presentations. The central viewpoint always occurred in each trial. The left viewpoint shows a rotation from center that retains the same visible parts, and the right viewpoint shows a rotation from center that shows different visible parts.

Figure 2. Experiment 1: Recognition (a) latencies and (b) errors for recognizing shaded images and silhouettes in sequential matching task. Note: in this figure, and all others, error bars show standard error of the mean.

Figure 3. Examples of objects used in Experiment 2 from “same” (first column) and “different” (second and third columns) groups, shown as (a) shaded images and (b) silhouettes.

Figure 4. Experiment 2: Facilitation in naming shaded images and silhouettes, when either the same object or a different object with the same name was presented previously.

Figure 5. (a) Stimuli used in Experiment 3, shown at 0°; (b) example stimulus showing 0°, 60°, and 180° viewpoints.

Figure 6. Recognition (a) latencies and (b) errors for recognizing shaded images in Experiment 3 across three depth rotations.

Figure 7. Naming latencies for shaded images presented in Experiment 4 at a 60° or 180° rotation from the initial viewpoint.

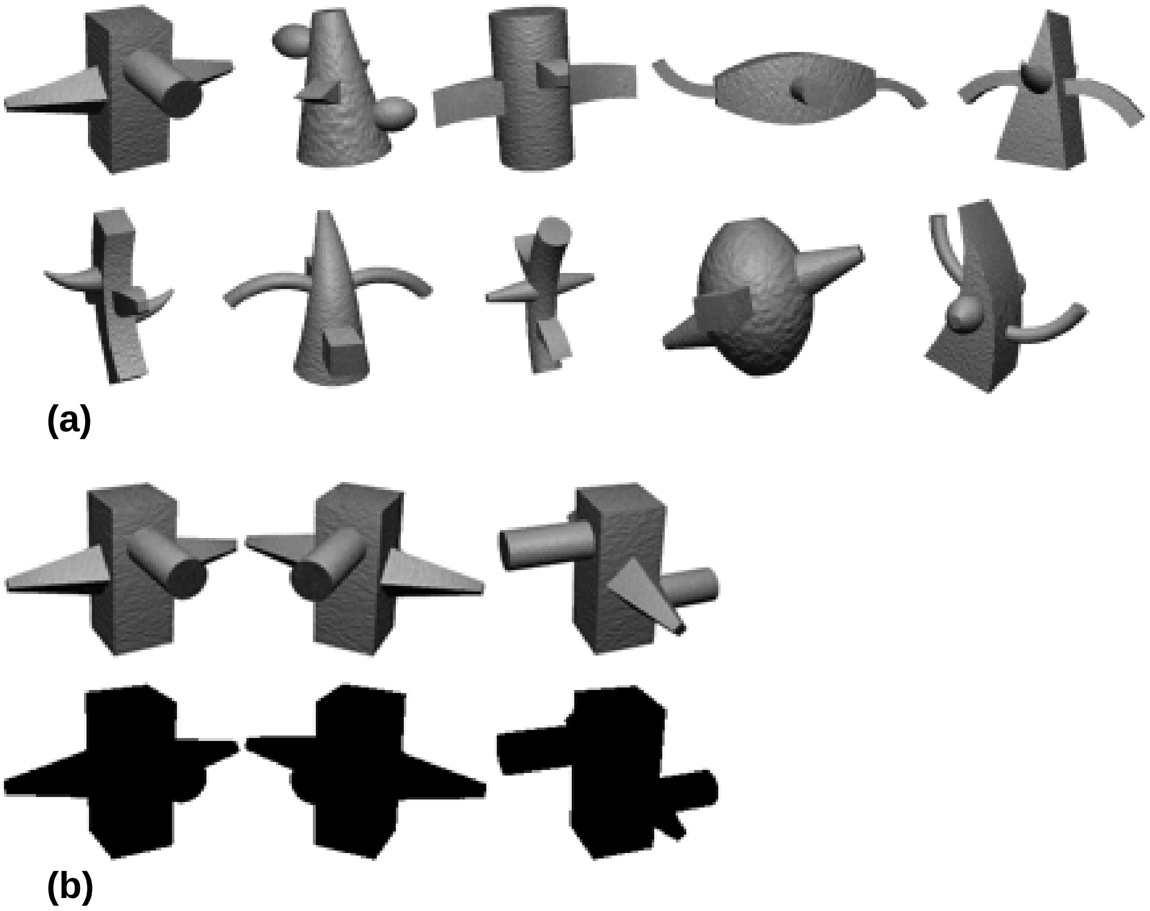

Figure 1

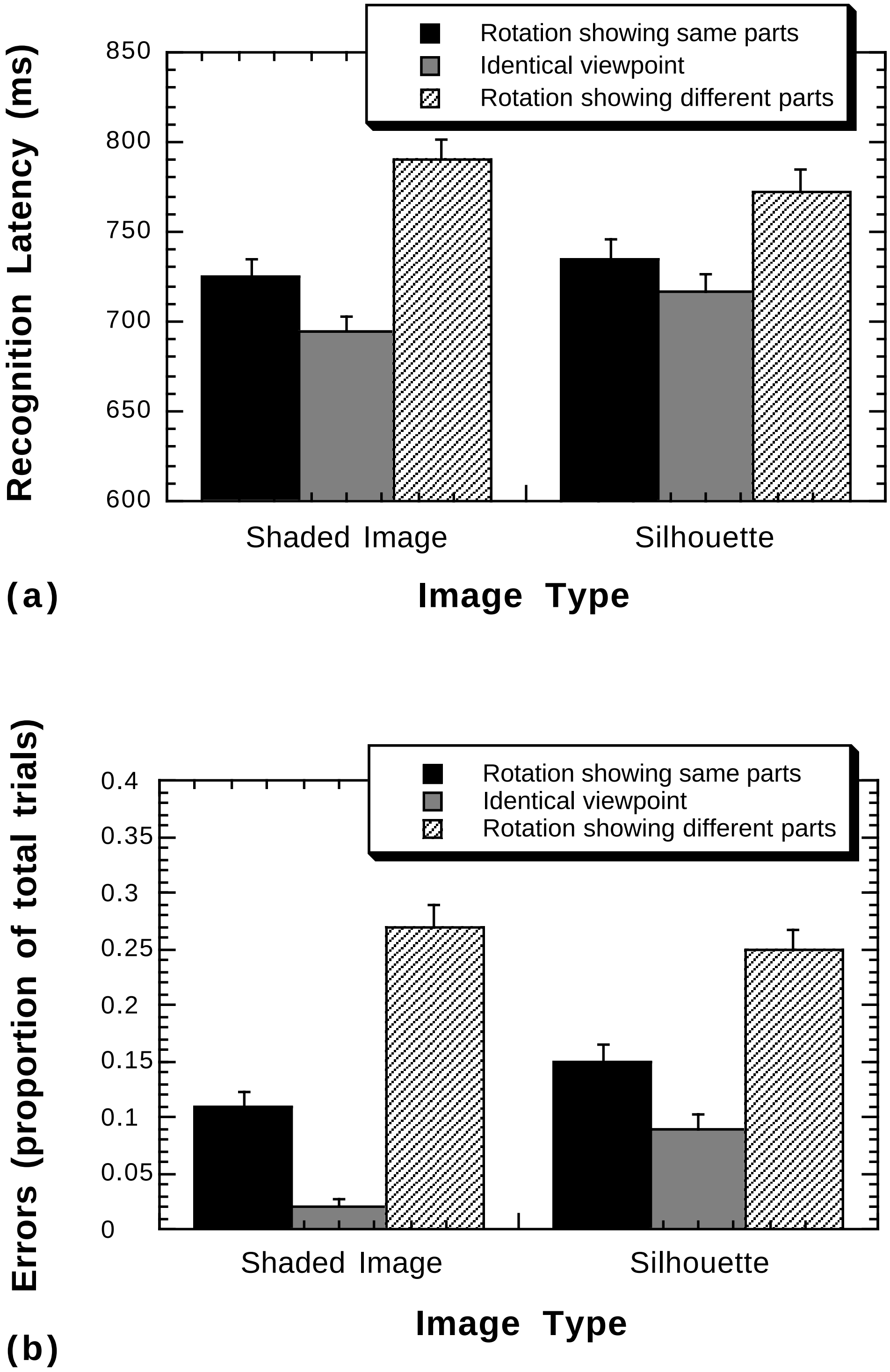
■ Rotation showing same parts
▨ Identical viewpoint
▨ Rotation showing different parts
Recognition Latency (ms)
850
800
750
700
650
600
Shaded Image
Silhouette
(a)
Image Type
■ Rotation showing same parts
▨ Identical viewpoint
▨ Rotation showing different parts
Errors (proportion of total trials)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Shaded Image
Silhouette
Image Type
(b)

Figure 2

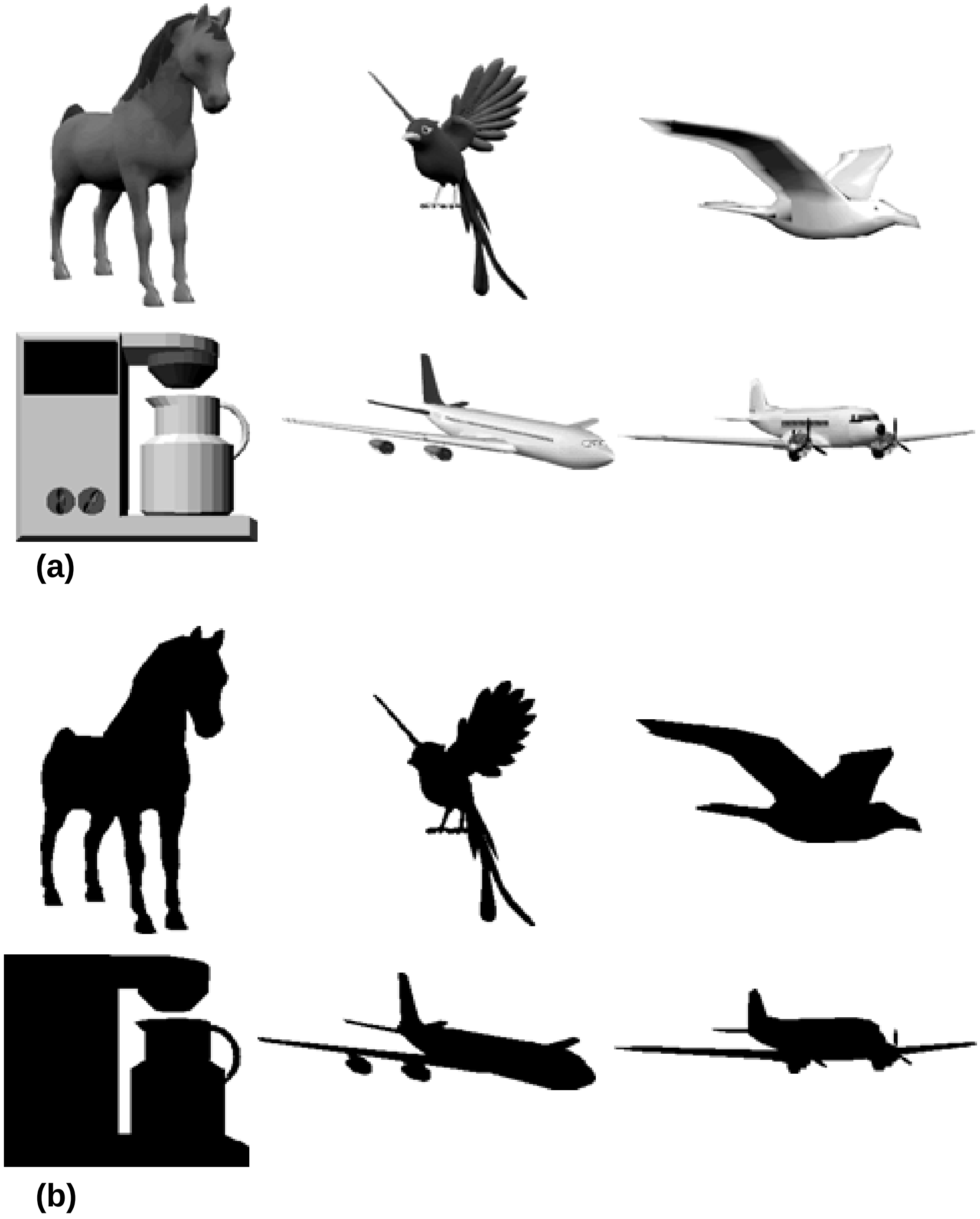

Figure 3

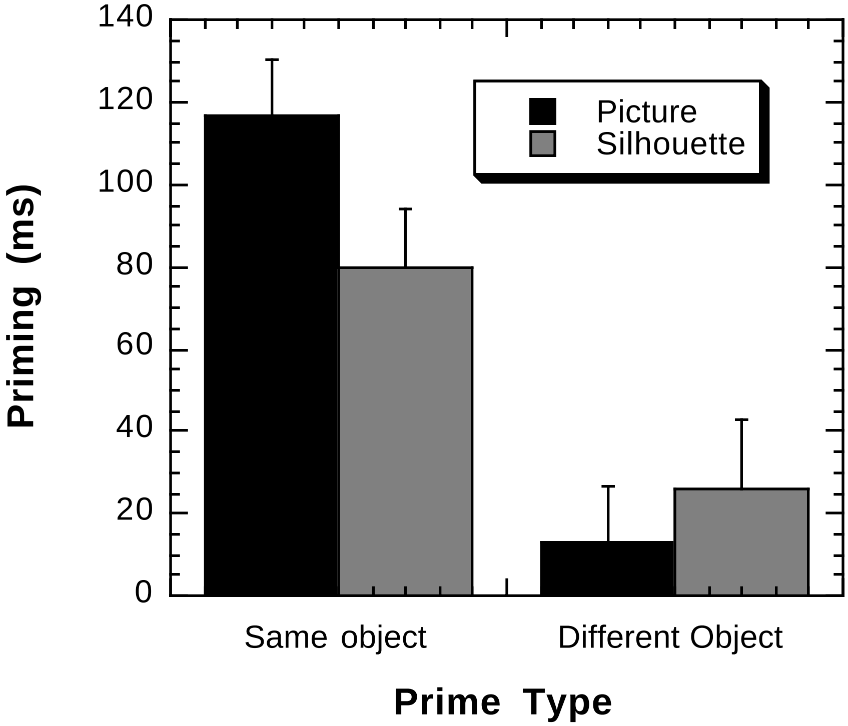
140
120
100
80
60
40
20
0
Priming (ms)
Picture
Silhouette
Same object
Different Object
Prime Type

Figure 4

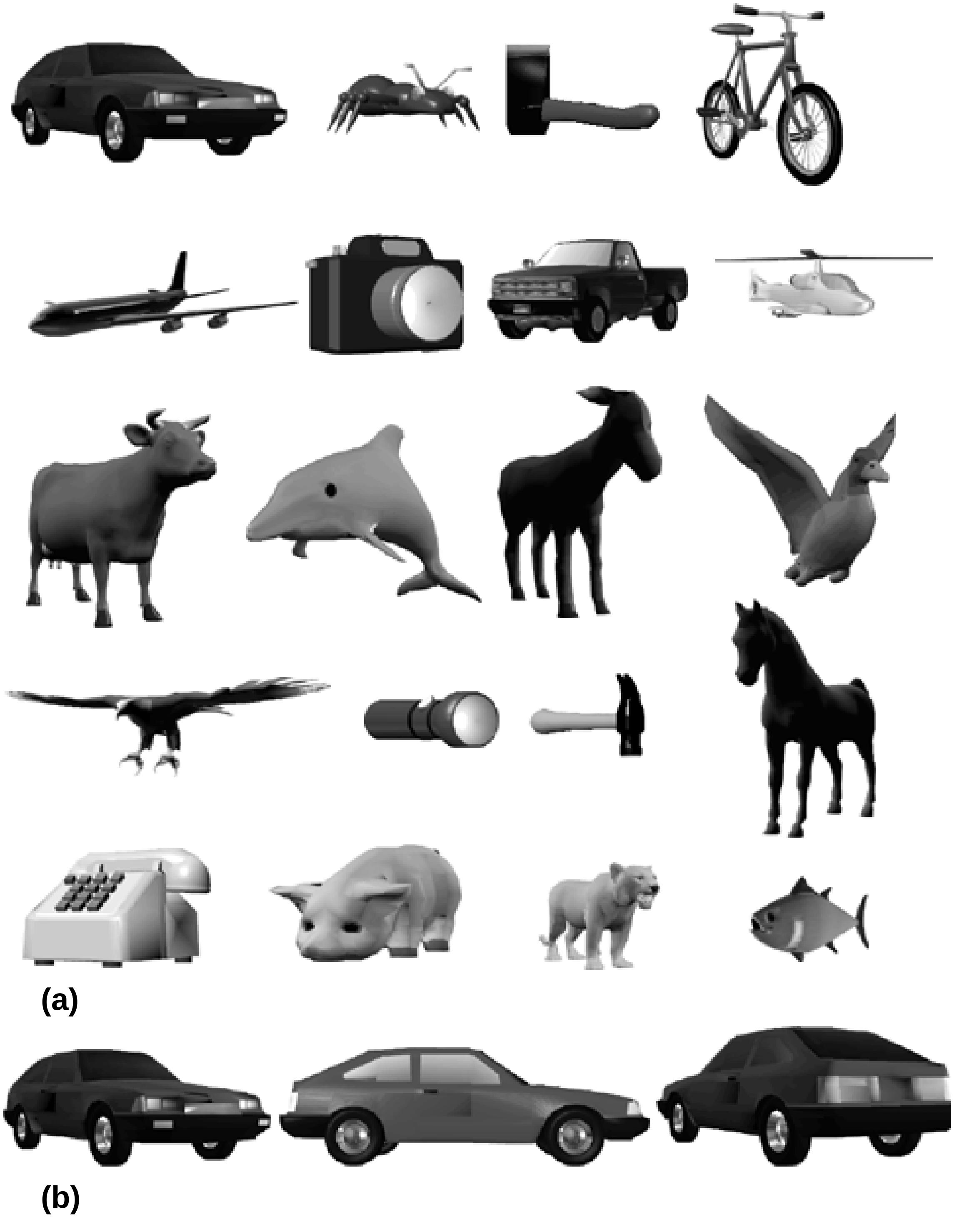

Figure 5

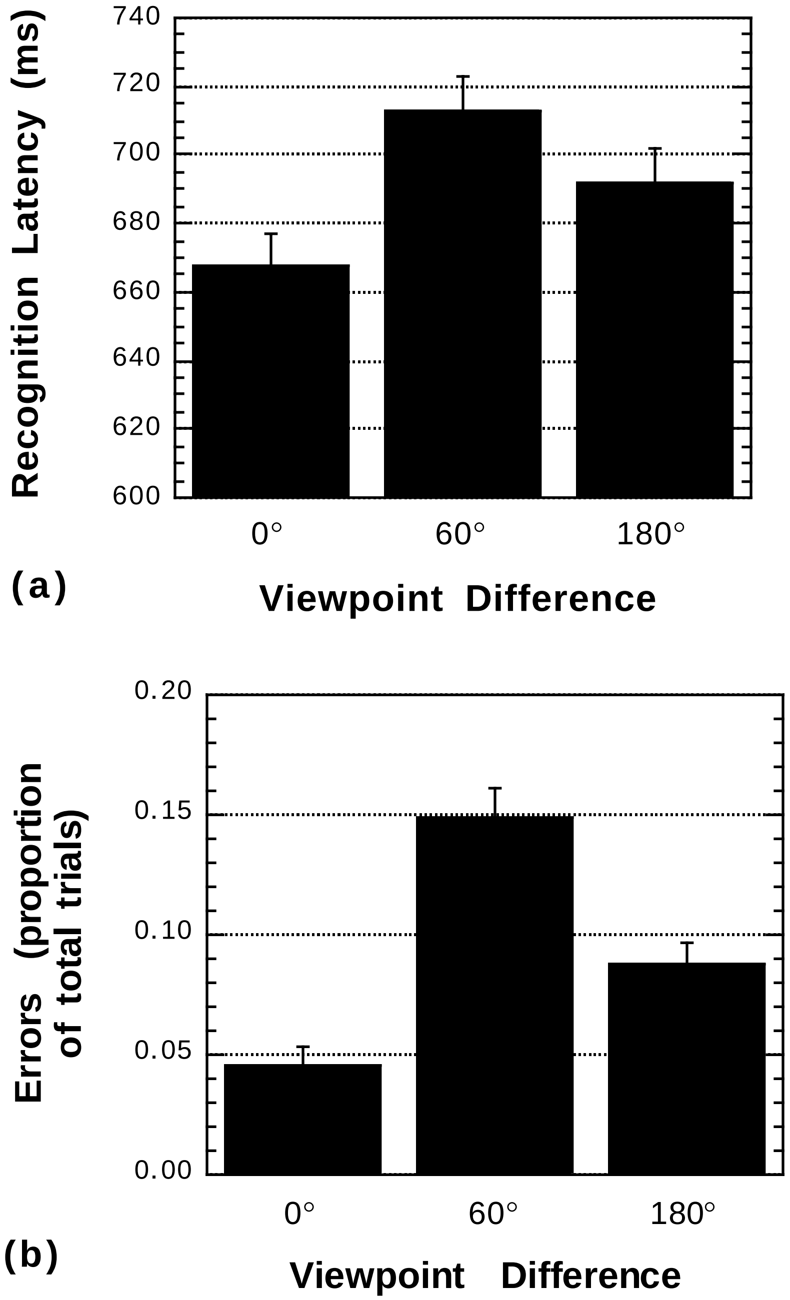
Recognition Latency (ms)
740
720
700
680
660
640
620
600
0°
60°
180°
(a)
Viewpoint Difference
Errors (proportion of total trials)
0.20
0.15
0.10
0.05
0.00
0°
60°
180°
(b)
Viewpoint Difference

Figure 6

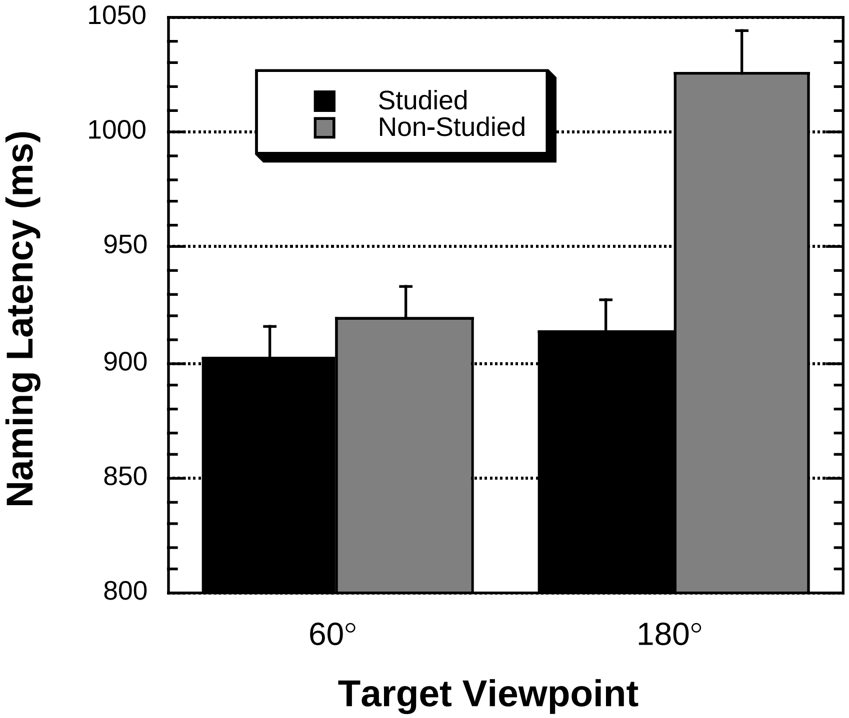
1050
1000
950
900
850
800
Naming Latency (ms)
■ Studied
■ Non-Studied
60°
180°
Target Viewpoint

Figure 7