
Recognizing silhouettes and shaded images across depth rotation

William G Hayward[#], Michael J Tarr[¶], Anna K Corderoy

Department of Psychology, University of Wollongong, Wollongong, NSW 2522, Australia; e-mail: whayward@psy.cuhk.edu.hk; [¶]Department of Cognitive and Linguistic Sciences, Brown University, PO Box 1978, Providence, RI 02912, USA;
Received 17 November 1998, in revised form 12 July 1999

Abstract. Outline-shape information may be particularly important in the recognition of depth-rotated objects because it provides a coarse shape description which gives first-pass information about the structure of an object. In four experiments, we compared recognition of silhouettes (showing only outline shape) with recognition of fully shaded images of objects, by means of a sequential-matching task. In experiments 1 and 2, the first stimulus was always a shaded image, and the second stimulus was either a shaded image or a silhouette. Recognition costs associated with a change in viewpoint were no greater for silhouettes than they were for shaded images. Experiments 3 and 4 replicated the design of the earlier experiments, but showed a silhouette as the initial stimulus, rather than a shaded image. In these cases, recognition costs associated with a change in viewpoint were greater for silhouettes than for shaded images. Combined, these results indicate that, while visual representations clearly include additional information, outline shape plays an important role in object recognition across depth rotation.

1 Introduction

Explaining how we recognize objects over changes in viewpoint is a fundamental problem for studies of human vision. Although early models of object recognition proposed representations that were based on the object, rather than the viewpoint of the observer (eg Binford 1971; Marr and Nishihara 1978), recent theoretical approaches to the problem of *view generalization* have stressed the importance of information derived from the observer's perspective (eg Biederman 1987; Hummel and Biederman 1992; Poggio and Edelman 1990; Tarr 1995; Ullman 1989). This theoretical development has been largely based on the experimental finding that when one view of an object is studied, not all other views are recognized with equal efficiency. Although some views are recognized as quickly and as accurately as the studied view, other views produce slower and more error-prone responses. This finding suggests that the information available at a particular viewpoint is of crucial importance. Current models of object recognition (eg Hummel and Biederman 1992; Poggio and Edelman 1990) propose that features are roughly encoded as they appear when an object is perceived. Such models predict that a cost in recognition (relative to recognizing the same view) will be obtained for most changes in viewpoint, because a compensation for the change in visual features will be necessary.

Despite general agreement that changes in viewpoint produce costs in recognition, there is an ongoing debate regarding the kinds of features upon which recognition is based. In an influential model, Hummel and Biederman (1992; following Biederman 1987) proposed that the crucial features used to recognize objects are volumetric [three-dimensional (3-D)] primitives that are fit to the visible parts of an object. If a view shows different parts (eg through self-occlusion), recognition costs are predicted to increase. Hummel and Biederman's model has difficulty accounting for the specific patterns of viewpoint dependence seen in a variety of behavioral studies (eg Hayward and Tarr 1997; Tarr et al 1997, 1998); moreover, their model has also received criticisms of its theoretical

[#]Present address: Department of Psychology, Chinese University of Hong Kong, Shatin, NT, Hong Kong

basis (Kurbat 1994; Tarr and Bülthoff 1995). These recent studies suggest that patterns of viewpoint dependence cannot be predicted accurately on the basis of visible parts.

Our ultimate goal is to elucidate exactly what types of features *can* be used to predict recognition costs over changes in viewpoint. Alternative models have attempted this to some extent [eg Bricolo et al (1998) propose local configurations of grey levels], but features are often chosen in an ad hoc manner, and are used to test a particular model rather than being proposed as a generic solution for object recognition. One major reason for this lack of precision in current models is the redundancy of visual information within objects. Any particular view of even a moderately complex object will contain a variety of surface, brightness, color, texture, and contour information, such that it is difficult to know where to start in specifying the particular subset of features that might mediate object recognition (if, indeed, such a subset exists).

The aim of this paper is to try to specify at least some of the visual features that are used in the process of view generalization. As noted, one difficulty in specifying the appropriate visual features is that typical views of objects contain vast amounts of information, all of which could conceivably be used to generalize from known to unknown views of an object. The aim of this study is to test recognition performance for a small subset of the possible visual features of an object; specifically, the features contained in the outline shape of an object (ie those features available from a silhouette). By taking a normal, shaded image and reducing it to a silhouette, we eliminate much of the available visual information, and are left with a small set of contours (which reflect both contours and surfaces of the object). If recognition costs for silhouettes are approximately the same as they are for shaded images, visual features in the outline shape of an object can be considered as viable candidate features for models of object recognition. If recognition of silhouettes is markedly more difficult than recognition of fully shaded images, one could conclude that nonsilhouette information, such as surface curvature and texture information, is crucially important for object recognition processes.

1.1 *Why examine outline shape?*

The restriction we impose (by focusing on outline shape) has a pragmatic justification; given the huge variety of possible sources of shape information in an object, the outline of an object will form a small subset of this information. However, there are also good theoretical reasons for investigating the utility of outline-shape information in object recognition over rotation in depth. In particular, the outline shape of an object will tend to show its major components, and will tend to be salient, given that it marks the boundary of figure from ground. Various computational models of shape representation are based upon silhouette information (eg Blum 1967; Kimia et al 1995; Richards and Hoffman 1985; Zhu and Yuille 1996), because it provides a computationally efficient method of segmenting the object (Zhu and Yuille 1996). Outline shape will also tend to be stable over rotations in depth, in that large components of an object will remain in the silhouette across relatively large changes in viewpoint.

Recently, we have shown that costs to object recognition following rotation in depth can be predicted, in some cases, by changes in outline shape. We (Hayward 1998) compared recognition, in a sequential-matching task, of shaded images and silhouettes following initial presentation of a shaded image, and found two principal effects. First, when the second stimulus showed the object in the same viewpoint as it had been presented initially, recognition was better for shaded images than silhouettes. Second, when the second stimulus showed the object rotated relative to the first, there was no difference in recognition performance between shaded images and silhouettes. This latter result suggests that outline shape may be particularly important for performing view generalization, when objects must be recognized across a rotation in depth between viewing instances.

We (Hayward and Tarr 1997) have also examined whether particular shape changes in the outline predict recognition costs following rotation in depth. All stimuli were fully shaded images of objects. In a sequential-matching task, the second stimulus might show an object from the viewpoint identical to that of the first stimulus, or from a viewpoint which was rotated 45° from the first stimulus. If rotated, the second stimulus might show an outline shape which was qualitatively different (Biederman 1987; Koenderink and van Doorn 1979) from the outline shape of the first stimulus, or the second stimulus might show an outline shape which differed only quantitatively from the first stimulus. A qualitative change in outline shape produced a significantly greater cost to recognition, in terms of both latencies and errors, than a quantitative change in outline shape.

Taken together, our results (Hayward 1998; Hayward and Tarr 1997) suggest an important role for outline shape in recognizing objects across rotations in depth. First, outline-shape information is often sufficient for recognition. Second, changing the qualitative information about silhouettes (eg Hayward and Tarr 1997) produces the same kind of effect on recognition performance as changing the qualitative information in shaded images [such as occluding a part (Biederman and Gerhardstein 1993)]. These investigations of outline shape, however, have failed to investigate the issue systematically [though see Newell and Findlay (1997) for a task using a range of viewpoints]. We previously drew conclusions about the importance of outline-shape information on the basis of discrete, pairwise comparisons with single sets of objects (Hayward 1998; Hayward and Tarr 1997). In order to understand the role of outline information, it is necessary to compare recognition of silhouettes and shaded images under somewhat more ecologically valid conditions, eg over a range of viewpoints. In addition, previous studies have only examined generalization from shaded images to silhouettes. If visual features in the outline provide the primary basis for view generalization, the size of the viewpoint costs for silhouettes should be similar to the size of the viewpoint costs for shaded images across parametric variation.

It is clear that an explanation of object recognition based upon outline shape will not form a complete theory of human performance; the visual system is able to readily discriminate a television from a microwave oven, even though each will have a similar qualitative outline shape. However, in this project we are restricting ourselves to one facet of the object-recognition problem: view generalization. Because of the stability of the shape information available in outlines, such information may be well placed to form the basis for view generalization from known to unknown views. Whether the visual features that are the basis for view generalization also provide the basis for recognition of individual exemplars of subordinate or entry-level categories is beyond the scope of the present study (and, indeed, our conjecture is that these sets of features are not identical).

1.2 *Overview of the experiments*

In this paper, four experiments are presented to test effects of viewpoint change on recognition in situations in which only outline-shape information is available. Recognition of shaded images and silhouettes is compared, following presentation of an initial stimulus. To ensure that the results generalize over different stimulus geometries, two different stimulus sets were used, presented in figure 1. The set used in experiment 1 [used by Hayward (1998), which in turn were based on stimuli used by Biederman and Gerhardstein (1993)], were qualitatively different arrangements of simple volumetric primitives. Each of these objects also had a large central component, but each component was a qualitatively different shape, which resulted in large variations in outline shape across the five objects. The set used in experiment 2 was designed so that all members of the set shared a basic spatial configuration; each object consisted of a central cylinder, with small additional components connected at different points in the surface of the

cylinder. Thus, the silhouettes of these objects were reasonably similar, differentiated by small changes in outline shape. Because of these differences between the objects, individuating features occurred in the outline shapes of most views of the objects in experiment 1. However, such features, if they occurred, were much smaller and less salient in the silhouettes of the objects in experiment 2. Thus, if the use of outline-shape information in object recognition is based upon the extent to which the outlines of the objects can be differentiated from one another, we should expect to see a greater reliance on outline shape in experiment 1. On the other hand, if outline shape is a property which is routinely exploited by the human visual system, performance on silhouettes in both experiments might be similar to performance on shaded images.

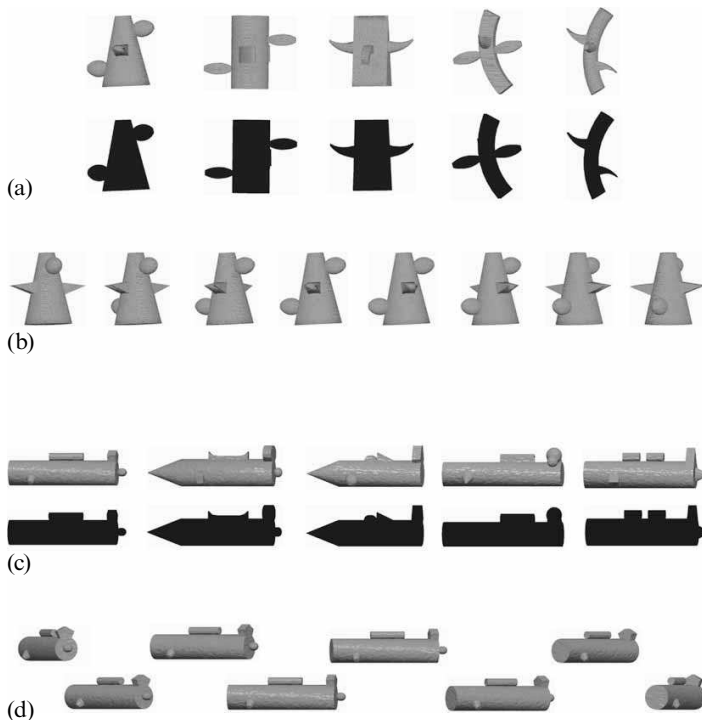


Figure 1. The objects used in the experiments: (a) objects used in experiments 1 and 3. These stimuli are based on sets used by Biederman and Gerhardstein (1993) and Hayward (1998). (b) All possible views of one object. (c) Objects used in experiments 2 and 4. (d) All possible views of one object.

In all experiments, a sequential-matching task was used. This task involved the presentation of one stimulus for 200 ms, a mask for 750 ms, the second stimulus for 200 ms, and a final mask. The task for the subject was to judge whether the second stimulus depicted an object that was the same as or different from the first stimulus. In all experiments, the second stimulus might be a shaded image of an object or a silhouette that showed only the outline shape of the object from a particular viewpoint. In terms of the processing required by this test, therefore, the task can be considered a 'working-memory' task, because to perform it correctly, participants must maintain a representation of the first stimulus. When the second stimulus is presented, the subject must 'compare' the perceived stimulus with the representation of the first stimulus in working memory.⁽¹⁾

⁽¹⁾In previous studies (eg Hayward 1998; Hayward and Tarr 1997; Tarr et al 1997), results using a sequential-matching paradigm have been found to be very similar to those obtained from long-term, naming studies. Thus, although we restrict ourselves here to one task, there is evidence to suggest that the results should generalize to other tasks.

The main function of experiments 1 and 2 was to test the extent to which generalization from one view of an object to another could be accounted for by changes in the outline shape of an object. In experiments 1 and 2, therefore, participants saw a shaded image of an object as the first stimulus. This presentation allowed the subject to encode the stimulus in the same way as in typical experiments [which show two-dimensional (2-D) depictions of apparently 3-D objects]. The second stimulus in these experiments could be either a shaded image or a silhouette. If outline-shape information is insufficient for object recognition across depth rotation, silhouettes should be more difficult to recognize as compared with shaded images, with the silhouette–shaded difference increasing across larger rotations of the object. Consider that a small rotation away from the studied viewpoint will produce a silhouette that has an outline close to the outline of the original image. Even if view generalization typically relies on internal-contour information, it would not be surprising if image-based perceptual processes were able to match the outline contour in one image with a very similar contour in the other image [eg via a ‘flexible template’ (Tarr and Bülthoff 1995)]. As the object rotates further, however, the outlines of the two images will become dissimilar, so a simple image match will not allow object identification to take place. Thus, this strategy for matching outlines will produce large costs at larger rotations, and costs that are larger than the costs for generalizing across views of shaded images. Such an effect will produce an interaction between the type of image and the size of the viewpoint change. On the other hand, a main effect of image type, without an interaction, suggests a general cost of changing the stimulus from a shaded image to a silhouette, but not one associated with generalizing from studied to novel views. This latter pattern of performance would suggest that the information contained in the outline shape of an object is sufficient for view generalization processes to operate as efficiently as they normally do, while at the same time hindering the operation of other aspects of the object-recognition process.

2 Experiment 1

We previously (Hayward 1998) used the objects from experiment 1 of this paper in a sequential-matching task, in which the first stimulus was always a shaded image and the second stimulus was either a shaded image or a silhouette. In that experiment, we found that the recognition of silhouettes of rotated objects was very similar to recognition of shaded depictions of rotated objects. It is difficult to generalize, however, from these results because only three viewpoints, and only one set of objects, were employed. Experiments 1 and 2 of the present paper were designed to provide a much more comprehensive investigation of the use of outline shape in recognition over depth rotation, by using a much larger range of viewpoints and rotations between stimuli, and by using two different sets of objects. In experiment 1, the first stimulus on each trial was always a shaded image of an object, and the second stimulus was either a shaded image or a silhouette. The second stimulus was also rotated by up to 80° relative to the first stimulus. On the basis of the earlier result (Hayward 1998), we expected to find similar costs to recognition following viewpoint change for shaded images and silhouettes.

2.1 Method

2.1.1 *Participants.* Thirty undergraduate students at the University of Wollongong participated in experiment 1 in order to fulfil a course requirement. Here and elsewhere, all participants were naïve as to the experimental hypotheses, and none participated in more than one experiment.

2.1.2 *Stimuli.* The objects used in experiment 1 were created on a Macintosh computer by means of 3-D modeling software (StrataVision 3D, Strata, St George, Utah). The objects are displayed in figure 1a. Each object had a qualitatively different volume as the central component, and two pairs of smaller volumes connected to the sides of the central

volume. Each of the objects had clear structural descriptions, as defined by geon theory (Biederman 1987; Biederman and Gerhardstein 1993, 1995; Hummel and Biederman 1992).

All objects were realistically rendered in orthographic projection by using an anti-aliased, ray-tracing algorithm at 20° increments rotated in depth around the center of the central component between 0° (arbitrarily designated as the front of the object) and 180° (arbitrarily designated as the back), noninclusive (see figure 1b for examples). Rotations for all objects were performed in the same direction. All rendered images were presented against a white background. Silhouettes were created for each viewpoint by rendering the background separately, which resulted in a black silhouette against a white background. Stimuli were presented in 8-bit color. The maximum dimensions of each stimulus were 358 pixels horizontally and 229 pixels vertically.

2.1.3 Design and procedure. Participants were informed that two objects would appear in quick succession and that they should quickly decide whether the two objects were the same or different. They were told that the objects might be presented from different viewpoints, and that the second image might be a shaded image or a silhouette, but that recognition decisions should be made solely on the basis of object identity regardless of any change in appearance between the two presentations. Each trial began with a fixation cross for 500 ms, followed by the first object for 200 ms, a 750 ms mask (a repetitive pattern derived from features of the objects in the experimental set), the second object for 200 ms, and then the same mask again which remained on the screen until a response was made. Response latencies were recorded from the onset of the second object, and a response deadline of 1500 ms was imposed. Trials in which participants did not respond by this limit were discarded. Participants were instructed to press the 'z' key on the keyboard (which was labelled 'SAME') if the two presentations showed the same object, and the 'm' key (labelled 'DIFF') if the presentations showed different objects.

Each of the five objects in each set appeared in 40 'same' trials and 40 'different' trials, for a total of 400 trials per participant. Viewpoint separation between presentations (for 'same' trials) was either 0° (the identical viewpoint), 20°, 40°, 60°, or 80°. There were 4 trials at each separation for each object. The 0°-separation pairs consisted of repetitions of the 20° viewpoint (ie a rotation of the object 20° from front was repeated at each presentation), the 40° viewpoint, the 60° viewpoint, and the 80° viewpoint. The 20°-separation pairs consisted of the following viewpoint pairings: 20°–40°, 40°–60°, 60°–80°, 80°–100°. This pattern was followed for all separations, so that the 80° separations consisted of the following viewpoint pairs: 20°–100°, 40°–120°, 60°–140°, 80°–160°. At each viewpoint separation, presentation order of the pair was varied across objects; for example, in the 20°-separation pair of 60°–80°, some objects were presented with the 60° view followed by the 80° views, while other objects were presented with the 80° view followed by the 60° view. Two versions of each 'same' trial were then generated; one in which the second stimulus was a shaded image, and one in which the second stimulus was a silhouette.

'Different' trials were generated by pairing two different objects. These distractor objects were matched in pose to the viewpoints used for 'same' trials, producing as many distractors at each viewpoint separation as there were targets. Trial order was randomly determined for each participant. A computer-generated 'beep' sounded if a response was incorrect. Breaks occurred randomly throughout the experiment.

2.2 Results

Trials in which participants did not respond within 1500 ms were terminated automatically during the experiment, and such trials were excluded from the analysis in this experiment and the other experiments reported in this paper. In experiment 1, this procedure resulted in the omission of 0.87% of trials in which the same object was shown in both presentations ('target' trials). The mean response latencies for correct 'target' trials and error rates for 'target' trials in experiment 1 are shown in figure 2.

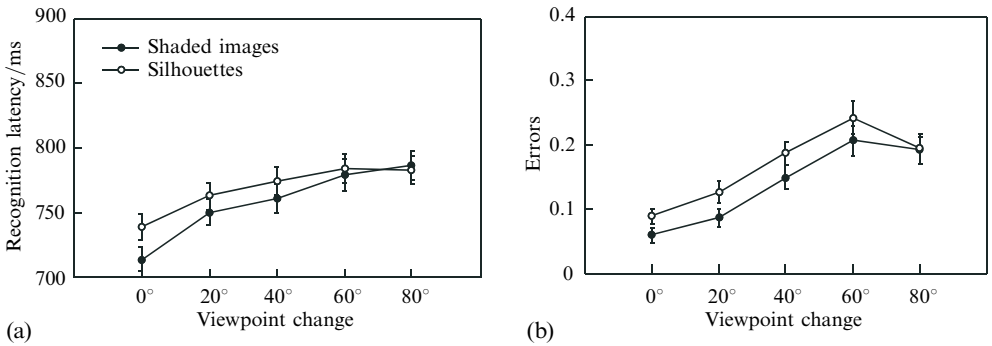


Figure 2. (a) Recognition latencies and (b) proportion of errors for recognizing shaded images and silhouettes in experiment 1. Error bars, here and elsewhere, show the standard error of the mean.

There appears to be a clear effect of viewpoint separation in these data, with recognition performance falling with increased separation between presentations. Analyses of variance (ANOVA) were performed on both the response-latency and error-rate data, with image type and viewpoint separation as within-subjects variables. For response latencies, there was a reliable main effect of viewpoint separation ($F_{4,116} = 13.08$, $p < 0.001$). The main effect of image type was not significant ($F_{1,29} = 3.07$, $p = 0.095$), nor was the interaction between viewpoint separation and image type ($F < 1$). For error rates, both main effects were statistically significant (for viewpoint separation, $F_{4,116} = 33.72$, $p < 0.001$; for image type, $F_{1,29} = 5.84$, $p < 0.05$), but the interaction was not reliable ($F < 1$).

To ensure that the effects of this analysis occurred evenly across the object set, and were not driven by a small set of items, analyses of items were computed, with the objects as the random factor, rather than subjects. For response latencies, viewpoint change was the only statistically reliable factor ($F_{4,16} = 16.18$, $p < 0.001$). Both the main effect of image type ($F_{1,4} = 2.3$, $p < 0.05$) and the interaction ($F < 1$) were non-significant. For errors, the same pattern was observed; viewpoint change was a reliable main effect ($F_{4,16} = 11.5$, $p < 0.001$), but image type ($F_{1,4} = 4.12$, $p = 0.11$) and the interaction ($F < 1$) were again nonsignificant. The failure to find a significant main effect in errors for image type contrasts with the reliable main effect for the same factor in the analysis by subjects. As image type is not a significant factor in the analysis of errors computed over items, it is likely that the main effect in the analysis over subjects does not occur evenly across the entire object set. Thus, conclusions regarding the generality of the image type difference in errors must be guarded.

As a means of ensuring that differences in the error rates were not due to manipulations in response criteria, we calculated a measure of sensitivity, $A',^{(2)}$ for all conditions of the experiment. Means are shown in table 1. We performed an ANOVA on these data, and found significant main effects for image type ($F_{1,29} = 22.0$, $p < 0.001$) and viewpoint difference ($F_{4,116} = 26.13$, $p < 0.001$). The interaction between these factors was also significant ($F_{4,116} = 2.9$, $p < 0.05$). Thus, unlike the error data, the sensitivity data show a significant interaction between image type and viewpoint separation, showing that the recognition costs over changes in viewpoint are reliably larger for silhouettes than shaded images. However, this interaction may be due to a ceiling effect, in that sensitivity in small viewpoint differences is very high in both conditions, and so performance in these cells may have been compressed. Reasons for this interaction will be considered in more detail below.

⁽²⁾ A' is a measure of sensitivity computed by means of a nonparametric model of signal detection, described by Donaldson (1992). Sensitivity varies from 1.0 (perfect performance) to 0.0 (perfectly imperfect performance), with chance performance at 0.5.

Table 1. Sensitivities (A') across conditions in experiment 1.

Viewpoint separation	Image type	
	shaded images	silhouettes
0°	0.94	0.94
20°	0.95	0.93
40°	0.92	0.91
60°	0.91	0.86
80°	0.92	0.88

2.3 Discussion

There were three main results in experiment 1. First, there was an effect of viewpoint change in both response latencies and error rates; as the viewpoint difference between the first and second image increased, recognition performance systematically became worse. This result suggests that recognition across a change in viewpoint was based on processes that were sensitive to changes in the specific appearance of an object. Second, there was mixed evidence as to the effect of a change of the stimulus from a shaded image to a silhouette. There was a small but significant effect of image type in the error data when analyzed over subjects, showing overall a greater proportion of errors for silhouettes than shaded images. The same result was obtained in the sensitivity data, but was not significant in any of the other analyses. Although the analyses are perhaps not conclusive on this issue, a result that performance was superior for shaded images compared with silhouettes would not be very surprising, given the huge change to the image (all color pixels being turned to black). At issue was whether such a performance deficit was related to the extent of the rotation between the two presentations. If outline shape is able to provide the basis for judgments of view generalization, we would expect that any small cost for silhouettes would be a constant over a change in viewpoint. Of interest, then, is the statistical test of the interaction.

As the interaction between image type and viewpoint change is statistically significant in analyses of the sensitivity data, but nonsignificant in analyses of both recognition latencies and errors for subjects and items, it is difficult to draw conclusions about the differential viewpoint costs for shaded images and silhouettes. As noted above, one possible reason for the interaction in the sensitivity data is a ceiling effect. To investigate these issues further, experiment 2 used the same paradigm as experiment 1, but with a set of objects that were qualitatively similar to one another. As such, this should reduce overall performance, and mitigate any ceiling effects.

3 Experiment 2

The stimuli from experiment 1 were chosen so that there were large qualitative differences in the shape of their components, particularly their central components, and so their silhouettes also contained large qualitative differences. Given these differences, it is possible that participants were able to perform the task in experiment 1 on the basis of selecting a few features that usually appeared in the outline of the object, regardless of whether the second stimulus was a shaded image or a silhouette.⁽³⁾ These large differences may also have contributed to a ceiling effect in the sensitivity data.

⁽³⁾In fact, two pieces of evidence suggest that this strategy was not employed by subjects in experiment 1. First, the linear nature of the viewpoint-cost function suggests the implication of a mechanism judging shape similarity in terms other than simply 'present' or 'absent'. Second, the fact that at the studied viewpoint subjects performed more accurately for shaded images (the studied stimulus) than silhouettes (the transformed stimulus) suggests again that their performance was at least influenced by nonoutline information (however, that nonoutline information may not have been employed in generalizing across viewpoint).

In experiment 2, objects consisted of a central cylinder, with small components connected at various points.

The issue of how to measure stimulus similarity is an important but complex one in computational vision, and has prompted much recent discussion (eg Cutzu and Tarr 1997; Edelman 1995, 1998; Hayward and Williams 1999). We have manipulated similarity in an obvious way in this experiment in order to reduce recognition-accuracy rates, and ensure the generalizability of our results. As such, we do not make theoretical claims regarding the effects of changing stimulus similarity; indeed, we wish simply to examine the extent to which *similar* results occur with different stimulus sets. However, we can present three pieces of evidence with which to support our claim that the stimuli used in experiment 2 were more similar to one another than those of experiment 1. First, we took silhouette versions of the stimuli from the viewpoint shown in figure 1a and did a pairwise calculation of the pixels that differed between the two images. The objects used in experiment 1 differed from one another by an average of 12% of the pixels in each image; the objects from experiment 2 differed by only 6.6% of pixels on average. Second, Biederman's (1987) recognition-by-components theory judges the stimuli from experiment 1 as having clearly different geon structural descriptions (as they were used by Biederman and Gerhardstein 1993), whereas the objects of experiment 2 have the same central component, and small parts which are not necessarily differentiable in terms of geon theory. Thus, although the objects of experiment 2 may have distinct geon descriptions (Biederman and Gerhardstein 1995), those descriptions will contain more similar geons to one another than descriptions of the experiment 1 stimuli. Last, and perhaps most importantly, error rates were lower and sensitivity rates were higher for experiments 1 and 3 than for experiments 2 and 4, showing that participants did have more difficulty discriminating the stimuli used in the latter experiments.

3.1 Method

3.1.1 *Participants.* Thirty-one undergraduate students at Brown University participated in experiment 2.

3.1.2 *Stimuli.* The objects were created on a Macintosh computer by using 3-D modelling software (StrataVision 3D, Strata, St George, Utah), and are displayed in figure 1. They were created by using a cylinder as the central component of each object. Small components were added at different spatial locations around each cylinder. The maximum dimensions of each stimulus were 358 pixels horizontally and 229 pixels vertically. Other attributes of the stimuli were the same as those in experiment 1.

3.1.3 *Design and procedure.* The experimental design and procedure was identical to that of experiment 1.

3.2 Results and discussion

In experiment 2, 1.98% of target trials were omitted because participants did not respond within the 1500 ms response deadline. The mean response latencies for the remaining correct 'same' trials and error rates for 'same' trials are shown in figure 3. For both response latencies and error rates, recognition performance was impaired as viewpoint separation increased. For latencies, there appeared little difference in performance between shaded images and silhouettes, except for trials on which the same viewpoint was shown on both trials, in which recognition of shaded images was faster than recognition of silhouettes. For errors, it appeared that performance was more accurate for shaded images throughout the trials.

ANOVAs were performed on both the response-latency and error-rate data, with viewpoint separation and image type (of the second presentation) as variables. For response latencies, there was a reliable effect of viewpoint separation ($F_{4,120} = 11.86$, $p < 0.01$) and a reliable interaction between viewpoint separation and image type

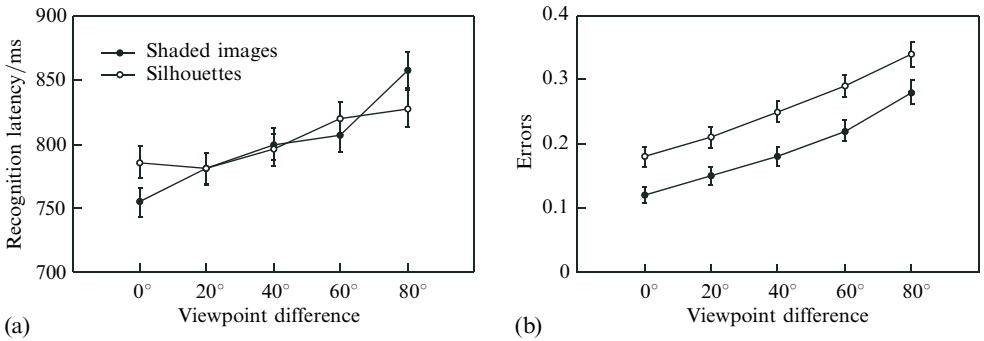


Figure 3. (a) Recognition latencies and (b) proportion of errors for recognizing shaded images and silhouettes in experiment 2.

($F_{4,120} = 2.84$, $p < 0.05$), showing, in this case, a *smaller* effect of viewpoint change on silhouettes than on shaded images. There was no main effect for image type ($F < 1$). For error rates, there were reliable main effects for both viewpoint separation ($F_{4,120} = 28.71$, $p < 0.001$) and image type ($F_{1,30} = 23.89$, $p < 0.001$). The interaction was not statistically significant ($F < 1$).

Analyses of items were again performed, to ensure that the statistical results occurred evenly across the stimulus set. For response latencies, viewpoint change was again a statistically significant main effect ($F_{4,16} = 33.82$, $p < 0.001$), but neither the effect of image type ($F < 1$) nor the interaction ($F_{4,16} = 1.23$) were significant. For errors, both main effects were significant (for viewpoint difference, $F_{4,16} = 31.67$; for image type, $F_{1,4} = 121.66$; in both cases $p < 0.001$) but the interaction was not ($F < 1$).

Table 2 shows the sensitivity rates for experiment 2 (again calculated as A'). An analysis of variance of these data with subjects as the random variable shows reliable main effects for both image type ($F_{1,30} = 68.33$, $p < 0.001$) and viewpoint change ($F_{4,120} = 26.98$, $p < 0.001$). The interaction, however, was not significant ($F_{4,120} = 1.07$, $p > 0.05$).

Table 2. Sensitivities (A') across conditions in experiment 2.

Viewpoint separation	Image type	
	shaded images	silhouettes
0°	0.91	0.85
20°	0.89	0.85
40°	0.88	0.81
60°	0.83	0.80
80°	0.82	0.74

The results of experiment 2 are similar to those of experiment 1. The effect of viewpoint change was reliable across all statistical tests, showing that performance was impaired when the object was rotated between the two presentations. When the second presentation was a silhouette, performance was reliably less accurate and less sensitive than when the second presentation was a shaded image (though correct responses for silhouettes were not slower than for shaded images). Again, however, any cost for recognizing silhouettes was not associated with the degree of viewpoint change. The interaction between viewpoint change and image type was significant in only one statistical analysis, and there it showed up as a *smaller* differential cost over viewpoint change for silhouettes as compared with shaded images. Thus, in no analysis was there any indication that silhouettes became differentially more difficult to match to the initial studied stimulus as the object was rotated away from its original viewpoint.

4 Discussion of experiments 1 and 2

Taken together, experiments 1 and 2 suggest that the recognition of silhouettes of rotated objects is quite close (in behavioral terms) to recognition of shaded images of rotated objects. In almost all conditions, the viewpoint costs for silhouettes were not statistically different from viewpoint costs for shaded images. The only analysis which showed a greater viewpoint cost for silhouettes was that for the sensitivity data of experiment 1. In experiment 2, however, when participants found the task more difficult, this interaction is no longer significant, suggesting the initial result is likely due to a ceiling effect having a strong influence on data in some cells. Certainly, there is no systematic pattern of results to suggest that silhouettes become differentially more difficult to recognize over rotations in depth.

These results suggest that, at the very least, visual processes which enable objects to be recognized from novel viewpoints are able to operate on outline-shape information about as efficiently as they *normally* operate on visual information from a fully depicted object. In other words, these experiments show that viewpoint generalization *can* occur on the basis of outline shape, and that the recognition-cost function that ensues (ie for silhouettes) will not be much different from the corresponding function for shaded images. What hypotheses can be formed, based upon the similar slopes of the recognition-cost functions for recognizing rotated shaded images and silhouettes? The most parsimonious, based on the experiments presented here as well as the results of similar studies (eg Hayward 1998; Hayward and Tarr 1997), is that view generalization is performed on the basis of outline-shape information alone. In all these experiments, recognition costs following the rotation of an object can be predicted on the basis of outline-shape differences between the studied viewpoint and rotated viewpoint of an object.

The hypothesis that outline-shape information is the basis upon which generalization to new views is performed would predict the results observed in experiments 1 and 2. Note that this hypothesis does not imply that outline-shape information is the only visual property to be encoded into an object representation. Much other information about objects can be attended to, and is likely to be crucial for decisions such as subordinate-level classification or differentiating between similarly shaped objects (eg fruit). However, the task used in the current experiments is a very specific one—albeit one that we frequently encounter in the natural world: judging whether a stimulus is the rotated version of one that was studied shortly beforehand. The hypothesis in question relates only to performance on this particular task.

To extend our hypothesis, experiments 3 and 4 were performed to investigate whether generalization to new viewpoints is based exclusively on outline-shape information. The experiments were identical to experiments 1 and 2 except for one variation: the first presentation of the stimulus was a silhouette rather than a shaded image. Thus, participants first saw a silhouette, and then were shown either another silhouette or a shaded image, and were asked whether the second stimulus depicted the same object as the first. In these experiments, as in experiments 1 and 2, all stimuli contained outline-shape information. If recognition of the rotated stimuli is primarily based on outline-shape information [eg if observers derive a purely contour-based description upon initial viewing, for instance, a ‘codon’ description (Richards and Hoffman 1985)], we expected to find no difference between recognition of shaded images and silhouettes, as each provides the same outline-shape information. Conversely, if generalization to new viewpoints involves nonoutline information, performance may differ between silhouettes and shaded images.

5 Experiment 3

5.1 Method

5.1.1 *Participants.* Thirty subjects from the University of Wollongong participated in experiment 3 for course credit.

5.1.2 *Design and procedure.* This experiment was identical in all respects to experiment 1, including the stimuli used, except that the first stimulus on each trial was always a silhouette. As before, the second stimulus was either a silhouette or a shaded image.

5.2 Results

In experiment 3, 0.87% of trials were excluded because participants failed to respond before the 1500 ms deadline. The mean response latencies for correct 'same' trials and error rates for 'same' trials are shown in figure 4. As in the other experiments, recognition performance, in terms of both response latencies and error rates, was impaired as viewpoint separation increased. For latencies, there appears little difference between the silhouettes and shaded images, except at the extremities; silhouettes were recognized a little faster than shaded images when there was no rotation, but a little slower when there was an 80° rotation between the stimuli. For errors, silhouettes show a much steeper function over degree of viewpoint change than do shaded images, suggesting that recognition of silhouettes was differentially worsened following rotation of an object.

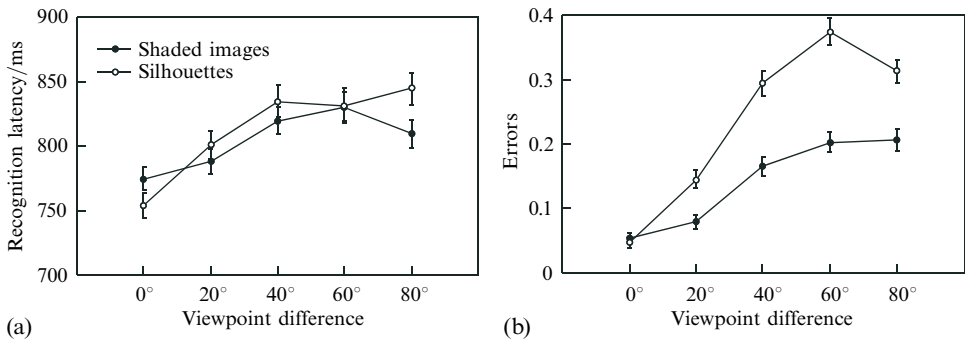


Figure 4. (a) Recognition latencies and (b) proportion of errors for recognizing shaded images and silhouettes in experiment 3.

Identical ANOVAs to those conducted in experiments 1 and 2 were performed on both the response-latency and error-rate data. For response latencies, there was a reliable main effect of viewpoint separation ($F_{4,116} = 22.36$, $p < 0.001$) and a marginally significant interaction between viewpoint separation and image type ($F_{4,116} = 2.38$, $p = 0.056$). The main effect of image type was not reliable ($F_{1,29} = 1.81$, $p > 0.05$). For error rates, both main effects were statistically significant (for viewpoint separation, $F_{4,116} = 91.53$, $p < 0.001$; for image type, $F_{1,29} = 67.65$, $p < 0.001$), as was the interaction ($F_{4,116} = 13.67$, $p < 0.001$).

Analyses of items were also computed. For response latencies, only the main effect of viewpoint change was statistically significant ($F_{4,16} = 21.29$, $p < 0.001$); the main effect of image type ($F_{1,4} = 2.81$, $p > 0.05$) and the interaction between viewpoint change and image type ($F_{4,16} = 1.79$, $p > 0.05$) were both nonsignificant. For errors, both main effects were significant (for viewpoint change, $F_{4,16} = 20.86$, $p < 0.001$; for image type, $F_{1,4} = 26.33$, $p < 0.01$), as was the interaction ($F_{4,16} = 7.93$, $p < 0.01$).

Sensitivity rates were also calculated, again by using A' , and are shown in table 3. An ANOVA calculated on these data showed a consistent pattern with those of the error-rate and recognition-latency data. There were main effects for viewpoint separation ($F_{4,116} = 65.83$, $p < 0.001$) and image type ($F_{1,29} = 13.78$, $p < 0.001$), and a significant interaction ($F_{4,116} = 4.64$, $p < 0.01$).

Table 3. Sensitivities (A') across conditions in experiment 3.

Viewpoint separation	Image type	
	shaded images	silhouettes
0°	0.93	0.95
20°	0.93	0.92
40°	0.90	0.89
60°	0.86	0.82
80°	0.88	0.85

5.3 Discussion

Following study of a silhouette, if the object was not rotated between trials, recognition of silhouettes was faster than recognition of shaded images.⁽⁴⁾ This result is not surprising, as it shows superior performance for recognizing the studied stimulus over an altered stimulus. However, if generalization from studied to novel views is performed on the basis of outline-shape information, recognition performance over changes in viewpoint should have been similar for silhouettes and shaded images, as each shares an identical amount of outline information with the originally studied stimulus. This was not the case. Accuracy (and sensitivity) decreased significantly more for silhouettes than for shaded images; the pattern is the same for response times, although the interactions are not significant. Thus, if a silhouette is studied as the initial stimulus, participants are more impaired if they see another silhouette than if they see a shaded image, as long as the object is rotated. This result suggests that, when generalizing from studied to novel views, outline shape is *not* the sole information used; rather, additional information, available in the shaded image, assists in generalization to the new views. Surprisingly, this additional information appears of less use if a shaded image is studied initially, because experiment 1 showed no difference in the recognition costs over depth rotation associated with shaded images and silhouettes. Nonoutline information becomes more useful if the outline was the *only* information available at the first presentation of the object.

6 Experiment 4

Experiment 4 was conducted to examine whether similar results would be forthcoming with objects which were more qualitatively similar than those used in experiment 3. Thus, experiment 4 was a replication of experiment 2, except that the initial stimulus in each trial was a silhouette instead of a shaded image.

6.1 Method

6.1.1 *Participants.* Twenty-nine participants from the University of Wollongong participated in experiment 4 in exchange for course credit.

6.1.2 *Design and procedure.* This experiment was identical in all respects to experiment 2, except that the first stimulus on each trial was always a silhouette. As before, the second stimulus was either a silhouette or a shaded image.

6.2 Results and discussion

As in the previous experiments, trials were automatically concluded if no response was made in 1500 ms. In this experiment, 1.24% of target trials were omitted for this reason. The mean response latencies for correct target trials and error rates for target trials are shown in figure 5. A rotation of the object between presentations appeared to

⁽⁴⁾Although recognition of shaded images was as accurate as recognition of silhouettes when the second viewpoint was identical to the initial viewpoint, this lack of difference was likely due to a ceiling effect, as accuracy was very high when there was no viewpoint change.

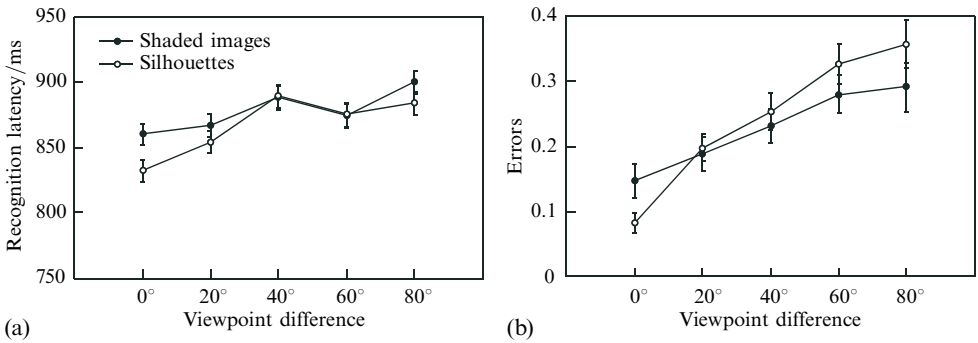


Figure 5. (a) Recognition latencies and (b) proportion of errors for recognizing shaded images and silhouettes in experiment 4.

produce differentially larger costs for recognizing silhouettes than shaded images. ANOVAs were performed on both the response-latency and error-rate data, with viewpoint separation and image type (of the second presentation) as variables. For response latencies there was a reliable effect of viewpoint separation ($F_{4,112} = 9.82, p < 0.001$). There was no main effect for image type ($F < 1$), nor was there a significant interaction between viewpoint separation and image type ($F_{4,111} = 1.37, p > 0.05$). For error rates, the main effect of viewpoint separation was reliable ($F_{4,112} = 28.92, p < 0.001$), as was the interaction between viewpoint separation and image type ($F_{4,112} = 3.9, p < 0.01$). The main effect of image type was not statistically significant ($F_{1,28} = 1.87, p > 0.05$).

Analyses of items were again performed, with objects as the random factor rather than subjects. For response latencies, the main effect of viewpoint separation was significant ($F_{4,16} = 19.04, p < 0.001$), as was the interaction between viewpoint change and image type ($F_{4,16} = 3.05, p < 0.05$). The main effect of image type was not statistically significant ($F_{1,4} = 2.92, p > 0.05$). For errors, the same pattern occurred: the main effect of viewpoint separation ($F_{4,16} = 45.99, p < 0.001$) and the interaction ($F_{4,16} = 4.16, p < 0.05$) were significant, but the main effect for image type was not significant ($F_{1,4} = 1.85, p > 0.05$).

As previously, sensitivity rates (calculated as A') were calculated (see table 4), and analyzed in an ANOVA. As in experiment 3, the results of this analysis were similar to the analyses of errors and recognition latencies. The main effect of viewpoint separation was statistically significant ($F_{4,112} = 13.72, p < 0.001$), as was the interaction ($F_{4,111} = 4.31, p < 0.01$). The main effect of image type was marginally significant ($F_{1,28} = 3.56, p = 0.07$).

Table 4. Sensitivities (A') across conditions in experiment 4.

Viewpoint separation	Image type	
	shaded images	silhouettes
0°	0.87	0.90
20°	0.83	0.86
40°	0.82	0.82
60°	0.83	0.81
80°	0.80	0.78

As in experiment 3, the results of the present experiment show an increase in the costs associated with recognizing silhouettes of rotated objects, given an initial silhouette, relative to recognition of shaded images. This increase in costs is statistically reliable in all analyses except for response latencies analyzed by subjects. The fundamental result of experiment 4, therefore, like experiment 3, is that changes in outline shape do

not predict performance on shaded images. The outline shape changed between the initial silhouette and a subsequent silhouette in exactly the same way that the outline changes between a silhouette and a shaded image, yet performance on the rotated silhouettes was impaired, relative to the shaded images. Performance in this experiment is not explained by appealing to a shape-generalization mechanism that operates exclusively on the outline shape of a stimulus. Surprisingly, the size of the interaction between silhouettes and shaded images across viewpoint changes was about the same here as in experiment 3. Thus, the silhouette-based processes we are studying do not appear to be sensitive to stimulus set homogeneity.

7 General discussion

This study was designed to test the extent to which changes in outline shape can account for recognition of objects across rotations in depth. Three general findings emerged, which held true across two sets of somewhat different objects. First, in situations when no rotation of an object occurred between study and test,⁽⁵⁾ recognition performance was best when the identical stimulus was repeated (a shaded image in experiments 1 and 2 and a silhouette in experiments 3 and 4). This result is not accounted for by the similarity in outline shape between the two stimuli, because the outline shape was obviously identical for either stimulus (shaded image or silhouette) shown at test. Rather, recognition of a stimulus from a repeated viewpoint, at least in the sequential-matching task, appears to be based on a computation of overall similarity between the two images. Any transformation of the stimulus impairs recognition performance to some degree. Because of this result, throughout the remainder of the paper we will consider the role of outline shape in view generalization, rather than simply in identifying a repeated pattern.

The second finding to emerge from this study relates the results of experiments 1 and 2. In these experiments, participants studied a shaded image and recognized either a shaded image or silhouette. When the object was rotated between study and test in these experiments, recognition performance was generally impaired, but the impairment was similar both for silhouettes and shaded images. Of the data analyzed, only one set showed a significantly greater recognition cost over changes in viewpoint for silhouettes than shaded images: the analysis of sensitivity in experiment 1. As noted, this result may have been caused by a ceiling effect, an interpretation which is supported by the fact that in experiment 2, when accuracy was reduced, the interaction was eliminated. In general, then, the viewpoint costs observed in experiments 1 and 2 were approximately equal for silhouettes and for shaded images.

The third finding of the study was that, in experiments 3 and 4, recognition of silhouettes no longer modeled recognition of the shaded images. In these experiments, analyses of sensitivities and hit rates always produced statistically reliable interactions, and in all cases the pattern of responses was compatible with a conclusion that recognition of silhouettes was more impaired by a change in viewpoint than recognition of shaded images. Given that the only information participants were able to encode from the studied stimulus was the outline shape of the target, one might have expected that the object would be recognized on the basis of outline shape. In that case, the outline of the test stimulus would have formed the basis for the recognition decision and, as the outline could be extracted from both the shaded image and the silhouettes, performance would be expected to be identical for both types of stimuli. The finding that recognition of silhouettes was less robust to a rotation in depth than recognition of shaded images shows that this type of explanation will not suffice.

⁽⁵⁾For the purposes of the rest of the paper the first presentation of each trial will be termed 'study', and the second presentation will be termed 'test'.

Why is recognition of depth-rotated objects predicted by changes in outline shape in some situations (experiments 1 and 2) but not in others (experiments 3 and 4)? This apparent paradox will be addressed below.

7.1 When outline shape predicts view generalization

In most object-recognition experiments, participants are presented with either line drawings or shaded depictions of objects. Each of these types of stimulus presents the participant with a wide variety of visual information about an object as it appears from the observer's viewpoint. Line drawings show both the bounding contour and internal contours, allowing recovery of the edges of elements of the object, which then allows recovery of the surfaces of the object. With shaded depictions the bounding and internal contours are again presented, but now surface-curvature, texture, and color information are also directly presented to the observer.

Structural-description theories of object recognition have tended to assume that edges are the fundamental building blocks of visual object recognition (eg Biederman 1987; Marr 1982). The results of experiments 1 and 2 provide a challenge to, although not necessarily a refutation of, such assumptions. When silhouettes were recognized as being particular objects, no information about internal contours/edges was available. Additionally, whereas some portions of the outline shape of the object show contours that are intrinsic to the shape of the object, such as the lip of a cup, other parts of the outline may be caused by the curvature of a surface (such as the sides of a cylinder), and so may be considered extrinsic (that is, the contour in the outline is not a property of a contour in the object). In a shaded image, intrinsic contours will generally represent boundaries between two surfaces, and therefore will always occur on the same location on an object if the object is rotated in depth. On the other hand, extrinsic contours will not occur on the same location of an object if it undergoes a rotation in depth (although similar extrinsic contours may occur across different views of an object). These contours provide different types of shape information about objects; intrinsic contours give the shape of contours, but extrinsic contours give the shape of surfaces. Recovering volume from outline shape will require use of intrinsic and extrinsic contours appropriately, an issue discussed in more depth by Tse (submitted).

Even if all elements of outline shape are treated as edges, the results of experiments 1 and 2 place constraints on the edges that are required for an object to be successfully recognized. We previously (Hayward 1998) argued that models of object recognition needed to be able to recognize objects from only outline-shape information. However, those conclusions were based on experiments using only a small number of object viewpoints, and in most cases used previously familiar objects. In this paper, we used novel stimuli and a wide range of viewpoints and, in experiments 1 and 2, viewpoint costs for the recognition of silhouettes were not generally different from recognition of shaded images. Any model of object recognition that *requires* the presence of internal contours or surfaces in a stimulus will fail to predict the results of these experiments.

7.2 When outline shape does not predict view generalization

As noted earlier, one possible implication of the results of experiments 1 and 2 is that view generalization can occur solely on the basis of outline shape. Certainly, in some situations this proposal must be true; if a cube is being discriminated from a sphere, it would not be surprising if this discrimination could be conducted on the basis of outline shape. If this proposal was generally valid, recognition should occur normally when only outline-shape information is available. The results of experiments 3 and 4 indicate that, when outline shape is all that is available at the point of study, recognition across depth rotation is improved when additional, nonoutline information is available in the test stimulus. These results clearly show that view generalization is impaired when outline shape is the only information available in visual memory.

Such a conclusion leads to a paradox of sorts; if recognition using only outline shape is impaired, then the results of experiments 1 and 2, showing equal performance for shaded images and silhouettes, need to be accounted for in terms other than of a simple match of outline shapes between the study and test stimuli. Equally, we need to account for recognition performance in experiments 3 and 4, when participants use more than just outline shape, even though in the study stimulus that is all that is available.

These two, apparently conflicting, requirements suggest that perceptual processes do not treat outline shape as a single, 2-D contour, because that contour is shown in both the silhouettes and shaded images in experiments 3 and 4, yet performance is different on these stimuli. If outline shape is not treated as a single, 2-D contour, how else could it be processed? Presumably, it is processed as visual information relating to the 3-D shape of an object. Although participants were presented with a silhouette in experiments 3 and 4, they knew that the silhouette represented a 3-D object. Indeed, whenever silhouettes are seen in the environment, they are known to be impoverished views of 3-D objects. Thus, it is likely that the representation of a silhouette is not the outline per se, but the object (or objects) that could be depicted by that silhouette. This analysis suggests that recognition of silhouettes might most closely follow recognition of shaded images when it is easiest to recover some aspects of the 3-D object structure from the silhouette.

Clearly, it would seem easier to recover the 3-D structure of an object from a shaded depiction than from a silhouette (eg by using shape from shading). Some ambiguity will remain, even with a shaded-image depiction, because the region behind the object will be occluded, and depth relationships between each point on the surface and the viewer can only be estimated from a static viewing (a problem alleviated, but not solved, by the addition of stereopsis). However, the ambiguity will be much smaller than that which happens when an attempt is made to recover a 3-D object from a silhouette. In this latter case, any change to the front of the object that does not affect the outline of the object will change the shaded depiction, but not the silhouette. On the other hand, any change to the object that affects the silhouette will also change the shaded depiction. Thus, if a set of possible target 3-D objects relate to any 2-D depiction, the set of relevant objects for a shaded image will be more constrained than the respective set for a silhouette.

In experiments 3 and 4, when participants studied silhouettes, recognition of silhouettes showed greater viewpoint costs than recognition of shaded images. When the first stimulus, a silhouette, was shown, that stimulus could represent a relatively unconstrained set of possible objects. Because only limited 3-D information can be computed from silhouettes, it is difficult to predict how the object it depicts is likely to appear from a new viewpoint. Consider what happens if the second stimulus is also a silhouette: if it is a rotated version of the first stimulus, neither image may contain sufficient information to easily determine the 3-D correspondence between the two 2-D images. On the other hand, if the second stimulus is a shaded image, information about the 3-D nature of the object may be sufficient to infer the appearance of that object at different viewpoints. Thus, the viewer may back-project this information to determine whether the initial silhouette is consistent with this shaded image (and vice versa).

In summary, the results of the experiments reported here suggest that outline-shape information is useful because it allows view-generalization processes to operate relatively efficiently, as long as some structural information is encoded about a target object. View generalization, however, does not proceed on the basis of outline shape alone, as demonstrated by the results of experiments 3 and 4. Rather, outline shape appears integrated into a richer object representation, and provides a powerful cue to activating that representation. The specific structure of the representation, and the manner in which it is activated by outline shape, is a topic for future research.

7.3 Conclusions

The results obtained in this study appear at first glance to be contradictory. We have shown that view generalization involving silhouettes shows a pattern of results similar to view generalization for shaded images, as long as the initial studied stimulus contains additional, nonoutline information. When the initial stimulus contains only outline information, however, recognition of silhouettes of rotated objects is impaired, relative to recognition of shaded images. These findings suggest that although changes in outline shape predict patterns of recognition performance across rotated objects, such information is but one element of a richer representation. In particular, shaded images allow more specific inferences about the 3-D shape of an object and such knowledge may facilitate better extrapolation regarding the appearance of outline shape from new viewpoints. What is still unknown is how both information about outline shape and internal surfaces and contours are encoded—that is, what specific features are used in the representation and how they interact during view generalization.

Acknowledgements. This research was funded by an Australian Research Council small grant to William Hayward, by a TRANSCOOP grant to Michael Tarr and Heinrich H Bülthoff, and by a Learning and Intelligent Systems Award IBN-9720320 from NSF to Michael Tarr. We would like to thank three anonymous reviewers for their comments. We would also like to thank Simone Keane and Stuart Johnstone for assisting in data collection.

References

- Biederman I, 1987 “Recognition-by-components: A theory of human image understanding” *Psychological Review* **94** 115–147
- 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** 1506–1521
- Binford T O, 1971 “Visual perception by computer”, paper presented at *IEEE Conference on Systems and Control* December, Miami, FL
- Blum H, 1967 “A transformation for extracting new descriptors of shape”, in *Models for the Perception of Speech and Visual Form* Ed. W Wathen-Dunn (Cambridge, MA: MIT Press) pp 362–380
- Bricolo E, Poggio T, Logothetis N K, 1997 “3D object recognition: A model of view-tuned neurons”, in *Advances in Neural Information Processing Systems* volume 9, Eds M C Mozer, M I Jordan, T Petsche (Cambridge, MA: MIT Press)
- Cutzu F, Tarr M J, 1997 *The Representation of Three-Dimensional Object Similarity in Human Vision* SPIE Proceedings from Electronic Imaging: Human Vision and Electronic Imaging II, San Jose, CA
- Donaldson W, 1992 “Measuring recognition memory” *Journal of Experimental Psychology: General* **121** 275–277
- Edelman S, 1995 “Representation, similarity, and the chorus of prototypes” *Minds and Machines* **5** 45–68
- Edelman S, 1998 “Representation is representation of similarities” *Behavior and Brain Sciences* **21** 449–498
- Hayward W G, 1998 “Effects of outline shape in object recognition” *Journal of Experimental Psychology: Human Perception and Performance* **24** 427–440
- Hayward W G, Tarr M J, 1997 “Testing conditions for viewpoint invariance in object recognition” *Journal of Experimental Psychology: Human Perception and Performance* **23** 1511–1521
- Hayward W G, Williams P, 1999 “Viewpoint dependence and object discriminability” *Psychological Science* in press
- Hummel J E, Biederman I, 1992 “Dynamic binding in a neural network for shape recognition” *Psychological Review* **99** 480–517
- Kimia B B, Tannenbaum A R, Zucker S W, 1995 “Shapes, shocks, and deformations, I: The components of shape and the reaction-diffusion space” *International Journal of Computer Vision* **15** 189–224
- Koenderink J J, Doorn A J van, 1979 “The internal representation of solid shape with respect to vision” *Biological Cybernetics* **32** 211–216

-
- Kurbat M A, 1994 "Structural description theories: Is RBC/JIM a general-purpose theory of human entry-level object recognition?" *Perception* **23** 1339–1368
- Marr D, 1982 *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (San Francisco, CA: W H Freeman)
- 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
- Newell F N, Findlay J M, 1997 "The effect of depth rotation on object identification" *Perception* **26** 1231–1258
- Poggio T, Edelman S, 1990 "A network that learns to recognize 3D objects" *Nature (London)* **343** 263–266
- Richards W A, Hoffman D D, 1985 "Codon constraints on closed 2D shapes" *Computer Vision Graphics and Image Processing* **32** 265–281
- 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, Bühlhoff 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, Bühlhoff H H, Zabinski M, Blanz V, 1997 "To what extent do unique parts influence recognition across changes in viewpoint?" *Psychological Science* **8** 282–289
- Tse P U (submitted) "A planar cut approach to volume recovery from silhouettes"
- Ullman S, 1989 "Aligning pictorial descriptions: An approach to object recognition" *Cognition* **32** 193–254
- Zhu S C, Yuille A L, 1996 "FORMS: A flexible object recognition and modeling system" *International Journal of Computer Vision* **20** 187–212

