Effects of Object Structure on Recognizing Novel Views of Three-Dimensional Objects

Scott H. Johnson
Department of Psychology
Cornell University
Uris Hall
Ithaca, NY 14853-7601
huc@cornella.cit.cornell.edu

Abstract

Can observers recognize novel views of three-dimensional (3-D) objects, created by rotations in depth from a single familiar view? Three experiments using 3-D model objects are reported demonstrating that: (a) subjects can indeed recognize novel views under these circumstances, and (b) recognition accuracy depends on the types of objects employed. More precisely, subjects successfully recognized geometrically regular and irregular objects rotated by 180 degrees about the vertical (y) axis. However, only geometrically regular objects were recognized when rotated similarly by 90 degrees.

These findings cannot be easily accommodated by contemporary object-centered or viewer-centered theories of shape-based object recognition, which make no provisions for representing different types of objects uniquely. Alternatively, these findings support a theory in which inferences about objects' 3-D shapes are generated from information implicit in their two-dimensional (2-D) bounding contours, or silhouettes (Johnson, 1993). Such inferences may be premised on rules that capture important regularities between 2-D bounding contours and 3-D surface geometry (e.g., Beusmans, Hoffman, & Bennett, 1987; Richards, Koenderink, & Hoffman, 1987).

Introduction

The ability to recognize visual objects based on their shapes presents a paradox. Any 3-D object can be viewed from an infinity of different viewpoints, each of which receives a unique image of that object in accordance with the laws of projective geometry. Consequently, recognition often demands identifying objects based on novel projections, or views. Virtually without exception, however, objects can be recognized immediately and effortlessly on the basis of their projected shapes. The question of how the visual system recognizes novel views of objects is of central importance to theories of object representation.

Supported by NIMH grant # 1F31MH10251-01 to SHJ.

Current Theories of Shape Representation

Of particular importance to recognizing novel views is the reference frame relative to which an object' size, location, and orientation are defined. Contemporary theories of shape-based recognition can be grouped into two families depending on whether the object representations they posit are specified relative to a frame of reference that is object- or viewer-centered.

Object-Centered Representations. Object-centered representations make explicit information about an object's entire 3-D structure, that is in no way dependent on a particular viewpoint (Marr & Nishihara, 1978). Object-centered representations are difficult to compute from an image, but once they are successfully recovered they can be matched against subsequently encountered views of the object, regardless of the observer's point of view (e.g., Huttenlocher, 1988; Lowe, 1987).

One of the greatest difficulties encountered by these theories concerns assigning an object-centered coordinate system to a projected image that is inherently viewpoint-dependent. What makes this task especially difficult is that each individual object must be consistently assigned its own frame of reference based on information available in the image, despite the fact that images projected by any one object can vary dramatically as a result of changes in viewpoint. Despite the existence of several different heuristics (e.g., Biederman, 1987; Marr & Nishihara, 1978), computing object-centered representations from viewpoint-dependent images remains a difficult task.

Viewer-Centered Representations. It has traditionally been assumed that the ability to recognize novel views is enabled by the existence of object-centered representations. Recently, however, this assumption has been questioned by theories employing multiple viewer-centered representations to depict an object's 3-D structure (e.g., Edelman & Weinshall, 1991; Perrett & Harries, 1988; Tarr, 1989; Seibert & Waxman, Forthcoming).

Models employing viewer-centered representations to depict 3-D shape have the advantage of eliminating what is perhaps the greatest difficulty with object-centered proposals: deriving viewpoint-independent descriptions from inherently viewpoint-dependent visual images. However, matching a viewer-centered representation with a description of a projected image often requires additional preprocessing to compensate for differences in viewpoints, and these processes must be implemented prior to determining the identity of the object. Furthermore, regardless of the compensatory strategy employed, recognition will be difficult whenever, resulting from effects of self-occlusion, the to-be-recognized image does not depict the same surfaces made explicit in at least one viewer-centered representation. Consequently, a very large number of viewer-centered representations might be required to represent all views of even a modestly complex object (Koenderink & van Doorn, 1976, 1979).

Recognizing Novel Views

The following experiments were undertaken to determine the coordinate system employed in representations of 3-D objects' shapes. In particular, two straightforward predictions arising from object- and viewer-centered theories were evaluated. If shape representations are object-centered, then subjects should be able to recognize novel views of an object based on knowledge of a single familiar view. However, if shape representations are viewer-centered, then subjects should have difficulty recognizing novel views based on familiarity with a single view; especially because the novel and familiar views tested depicted different surfaces.

General Method

Apparatus. A three-channel tachistoscope was modified to present 3-D model objects. A base, consisting of 1/2 inch diameter plastic pipe, was fitted inside the center of one chamber. Stimulus objects fit snugly over this base and could be manually interchanged through a door in the rear of the chamber, as well as rotated about the vertical, or y-axis. The height of the base prevented it from being seen through the subjects' window. A second chamber contained a white tachistoscope card with a central fixation "x" drawn in the center.

The tachistoscope was interfaced with an IBM PC model xt. The computer displayed the trial lists to the experimenter, and recorded subjects' responses from the keyboard.

Stimuli. Three different types of unfamiliar objects were used: (a) 16 objects constructed from plumbing fittings; (b) 16 closed wireframe objects, and (c) 16 objects modeled from clay. All of the objects used were asymmetrical about the axis of rotation (i.e., y or vertical axis). Construction of the different stimulus sets is detailed separately for each experiment below.

Procedure. The design was such that subjects only experienced each test item once from a single orientation and then were asked to recognize that item from either the same orientation, or a novel orientation created by rotating the object clockwise around the y-axis. Position on both the z- and x-axes was held constant throughout all experiments.

In each experiment, subjects were asked to study four objects from the total set of 16 items, in preparation for a memory test that would involve recognizing them from both novel and familiar views. Subjects were told that "objects in the subsequent recognition test will appear at both the orientation studied and at other orientations created by rotating them around the vertical axis." A model object, similar to those in the experiment, was used by the experimenter to demonstrate such rotations. At the beginning of each trial, subjects fixated a centrally presented "x." An object was then displayed for four seconds. The object was subsequently replaced by the fixation point and the object for the next trial was positioned by the experimenter. This sequence was repeated four times until the subject had studied each of the four target stimuli once.

Next, subjects performed a two-choice recognition task consisting of eight objects. Four were the previously studied target items, each appearing in one of four different orientations: the studied orientation, or rotated by 90, 180, or 270 degrees around the y-axis. The remaining four objects were previously unseen distracter items. Subjects were instructed to press a key labeled "yes" if they recognized the object as being one of those studied, regardless of whether or not it appeared in the previously trained orientation, or "no" if the item seemed unfamiliar. If subjects were uncertain as to whether the item was familiar, they were instructed not to deliberate, but to execute their first guess.

Order of presentation in the study session was counterbalanced across each group of eight subjects, and ordering of the test trials in the recognition phase was random for each subject. After a short break of approximately two minutes, the entire procedure was repeated for a second block with eight previously unseen objects: four new target objects and four new distracter objects.

For an individual subject, each of the 16 stimulus objects appeared only once in the recognition task, either as a target or as a distracter. Each of the four target orientations was, however, tested twice: once in each of the two blocks. Across every group of eight subjects, each of the 16 stimulus objects appeared once as a target in each of four test orientations.

Experiment One

The first experiment sought to determine whether subjects could recognize novel views of opaque objects created from geometrically regular parts, based on knowledge of a single familiar view. It was reasoned that if these objects were represented relative to an object-centered frame of reference, then all views should be recognized equally well regardless of whether they are familiar or not; That is, rotating objects by 90 or 180 degrees around the y-axis should have no effect on recognition accuracy. If, however, these objects were represented in a viewer-centered format, then recognition of novel views should be considerably less accurate than recognition of the familiar view.

Method

Subjects. Thirty-two undergraduate volunteers participated in a single testing session lasting approximately 15 minutes. Subjects received $2.00 for their participation.

Stimuli. Stimuli for experiments one through four were created by joining five 1/2" white plastic plumbing fittings together: a t-joint, two 90 degree joints, and two 45 degree joints. Use of identical parts insured that the objects could only be distinguished on the basis of their overall shape configuration, and not on the presence and/or absence of a distinctive part, overall differences in surface area, or volume. Parts were joined arbitrarily except that each object was asymmetrical and had a central t-joint oriented vertically. This made certain that the objects' principal axes were equally visible from each view, and provided a "trunk" onto which the remaining four parts were fixed. Increments of 90 degrees were marked around the base of the t-joint so that objects could be rotated to the desired test orientations when seated on the mount inside the tachistoscope chamber. The frame of the tachistoscope's chamber was adjusted so that the incremental markings were not visible to the subject.

Procedure. The procedure was as described in the General Methods section above.

Results and Discussion

Results from Experiment One indicate that subjects are indeed capable of recognizing novel views of 3-D objects based on knowledge of a single familiar view, as predicted by theories positing object-centered representations. There was no difference in hit rates between the 90 and 270 degree conditions [$t(31) = .97$, $p = .338$)], and these scores were combined to yield a single variable for 90^0 rotations independent of their direction. A one-way ANOVA performed across the 0, 90, and 180 degree conditions revealed no effect of orientation change on recognition accuracy ($F(2,125) = 2.53$, $p = .08$). More importantly, planned comparisons revealed no differences in accuracy of recognition between the familiar view and views rotated by either 90 ($t(31) = 1.03$, $p = .31$) or 180 degrees ($t(31) = 1.23$, $p = .23$) (Figure 1a).

Experiment Two

Results of Experiment One support the hypothesis that shape is represented in an object-centered format. Subjects were able to accurately recognize novel views of plumbing part objects based on knowledge of a single view. In contrast, a similar study by Rock et al. (1981) found a large drop in recognition accuracy when wireframe objects were rotated by 90 degrees around the y-axis, but detected no effect when the same objects were rotated similarly by 180 degrees. Why do observer's have trouble recognizing novel views of wireframe objects created by 90 degree rotations in depth, but no difficulty recognizing unfamiliar views of plumbing part objects? There are two possible explanations for these differences. First, the procedure used in the present experiments differs from that employed by Rock et al. (1981). Subjects in the present experiments were explicitly instructed to encode objects in a manner that would facilitate later recognition of novel views. In contrast, Rock et al. (1981) made no mention of the subsequent recognition task until after the encoding phase of the experiment was completed. Second, the stimulus objects used in the present experiments were quite different from the irregularly shaped wireframe objects used by Rock et al. (1981). Objects in Experiment One were opaque and constructed from several geometrically regular parts. Perhaps these structural differences affected the format in which shape was represented. Experiment Two sought to distinguish between these possibilities by testing subject's recognition of novel views of wireframe objects with the procedure used in Experiment One. If the differences between the present results and those of Rock et al. (1981) are attributable to procedure, then wireframe objects, like those constructed of plumbing parts, should be recognized in a viewpoint-independent manner. However, if the inconsistency between results are attributable to structural differences in the two object sets then, like Rock et al. (1981) reported, wireframe objects should be difficult to recognize when rotated by 90 degrees.

Method

Subjects. Twenty-four undergraduate volunteers participated in the 15 minute experiment for $2.00. Subjects were naive regarding the hypotheses under investigation, and had not participated in any related experiments.
Stimuli. Sixteen closed wire frame objects were created using 16 inch lengths of 2mm diameter steel wire. Lengths of wire were bent into closed arbitrary shapes, with both ends of the wire inserted into a plastic base that fit onto the mount inside the tachistoscope. This allowed each object to be precisely rotated to the desired orientation.
Procedure. With the exception of different stimuli, the procedure was identical to that employed in Experiment One.

Results and Discussion

Accuracy did not differ between 90^0 and 270^0 rotations, $t(23) = .87$, $p = .396$. Data from these orientations were therefore combined to yield a single value for 90^0 rotations independent of their direction. In contrast to data showing viewpoint-independent recognition of plumbing part objects, a one-way ANOVA revealed a significant effect of orientation change on recognition accuracy, $F(2,93) = 3.54$, $p = .033$. Planned comparisons evidenced a drop in recognition for views rotated by 90 degrees when compared with unrotated views, $t(23) = 2.23$, $p = .036$. In contrast, views rotated by 180 degrees were recognized as accurately as unrotated views, $t(23) = .7$, $p = .49$ (Figure 1b).

Results from Experiment Two were consistent with those of Rock et al. (1981), yet equivocal with regard to the frame of reference used in shape representations. As predicted by object-centered theories, novel views created by rotations of 180 degrees around the y-axis were accurately recognized. However, novel views rotated by 90 degrees were difficult to identify, as expected if representations of shape are viewer-centered.

Successful replication of Rock et al's. (1981) results suggests that differences in the ability to recognize novel views of wireframe objects, versus those constructed from plumbing parts, are attributable the way in which various types of shape information are represented in the visual system, rather than to procedural differences. This finding is of particular interest because current models of shape representation make no provisions for different types of objects; All objects are represented similarly.

Experiment Three

Ostensibly, plumbing part objects used in Experiment One differ from wireframe objects in at least two respects: (a) Wireframes have virtually no opaque surfaces, making it more difficult to perceive their 3-D structure. This may simply make it difficult to compute object-centered representations. (b) Wireframe objects also lack the geometrically regular structure found in plumbing part objects. Perhaps this particular difference affects the format in which shape is represented. Either of these differences might explain why subjects have difficulty recognizing wireframe objects rotated by 90 degrees in depth, but no trouble with similarly transformed plumbing part objects. A set of irregular clay forms was created to disambiguate these possibilities. On one hand, these clay objects are like plumbing parts in that they are: opaque, have good depth cues, and undergo substantial effects self-occlusion during changes in viewpoint. On the other hand, like wireframes, clay blobs lack regular geometric structure. If the presence of salient depth cues is critical for establishing object-centered representations, then recognition of clay blobs should be similar to that of plumbing part objects: There should be no effect of orientation change on recognition accuracy. Alternatively, if regular geometric structure is critical for establishing object-centered representations, then recognition of clay blobs should be similar to recognition of wireframe objects: There should be a drop in recognition accuracy for views rotated by 90 degrees.

Method

Subjects. Sixteen undergraduate students received $2.00 for their voluntarily participation in a 15 minute experiment. None of the subjects were familiar with the hypotheses being tested, or had taken part in other experiments reported.
Stimuli. Sixteen, asymmetrical objects were constructed from 2"x 2" x 4" blocks of gray modeling clay. The objects were formed into arbitrary 3-D shapes and fired in an oven. Plastic fittings were fixed to the bottom of the objects so they could be mounted to the base inside the tachistoscope and rotated to the desired orientation.

Procedure. The procedure was identical to that used in Experiments One and Two, except that 16 clay objects were substituted.

Results and Discussion

Unexpectedly, accuracy data from 90^0 and 270^0 rotations differed significantly [$t(15) = 2.52$, $p = .023$], and thus were kept separate for subsequent analyses. A single factor ANOVA revealed a highly significant effect of orientation change on recognition accuracy, $F(3, 60) = 7.143$, $p < .001$. Interestingly, this effect appeared attributable to a drop in recognition accuracy for objects rotated by 270^0, $t(15) = 3.87$, $p = .002$. Neither 90^0 [$t(15) = 1.0$, $p = .333$] nor 180^0 [$t(15) = .324$, $p = .751$] rotations had a significant effect on recognition accuracy.

There was no a priori reason to expect that 90 degree rotations in one direction should be any more difficult to recognize than equidistant rotations in the other direction. Nevertheless, subjects did find views rotated by 270 degrees more difficult to identify than views rotated by 90 degrees. Presumably this effect was attributable to unintended differences in the arbitrarily created clay shapes, that made views rotated 90 degrees in one direction more difficult to recognize than views rotated 90 degrees in the other direction. This fact, however, does not bear on the present issue of whether 90 degree rotations on the whole are more difficult to recognize than unrotated views. Therefore, performances on views rotated by either 90 or 270 degrees were averaged and submitted, along with data from views rotated by 180 degrees, to a oneway ANOVA. Consistent with the original analysis, there was a highly significant effect of orientation change on recognition accuracy, $F(2,61) = 5.88$, $p = .005$. A planned comparison revealed that views rotated by 90 degrees were less accurately recognized than unrotated views, $t(15) = 3.033$, $p = .008$ (Figure 1c).

Results from Experiment Three suggest that differences in the ability to recognize novel views of wireframe versus plumbing part objects were attributable to variations in object structure, rather than problems recovering depth information. Similar to results obtained with wireframe objects, subjects had considerable difficulty recognizing geometrically irregular clay objects rotated by 90 degrees, even though they have unambiguous depth cues. Likewise, rotations of 180 degrees did not affect recognition accuracy.

Summary and Conclusions

In short, the present experiments provide evidence that subjects can indeed recognize novel views of 3-D objects rotated in depth, based on knowledge of a single view. The extent of viewpoin independent recognition was, however, shown to va depending on the *type* of object tested. More precisely, subjects successfully recognized geometrically regular and irregular objects rotated b 180 degrees about the y-axis. However, only geometrically regular objects were recognized when rotated similarly by 90 degrees. These findings are difficult to account for in terms of contemporary theories that propose either object- or viewer-centere representations to represent all types of objects.

One-hundred eighty degree rotations in depth, that completely interchange visible and occluded surfaces, have no affect on the object's silhouette, other than reflecting it about the axis of rotation. In contrast, 90 degree rotations in depth only partly interchange visible and occluded surface but drastically affect the object's projectd silhouette. Interestngly, the former do not affect the accuracy o recognizing novel views, while the later may, depending on object's structural properties. Based o these and related observations, Johnson (1993) suggests an alternative theory of shape-based recognition wherein inferences about objects' 3-D shapes are generated based on information contained in their 2-D bounding contours, or silhouettes. It is claimed that these inferences are premised on rules that capture important regularities between projected 2-D bounding contours and 3-D surface geometry (e.g., Beusmans, Hoffman, & Bennett, 1987; Richards, Koenderink, & Hoffman, 1987). When information in the projected 2-D silhouette is sufficient to enable accurate inferences about the shape of occluded surfaces (e.g., the geometrically regular objects), all views of an object will yield the same 3-D representation, and recognition will be viewpoint-independent. However, when informatio in the silhouette is insufficient to unambiguously specify the shape of occluded surfaces (e.g., the geometrically irregular objects), 3-D interpretations preferred for views may differ, resulting in viewpoint-dependent recognition. Further work is currently underway to evaluate these hypotheses.

References

Beusmans, J. M. H., Hoffman, D. D., & Bennett, B. M. (1987). Description of solid shape and its inference from occluding contours. *Journal of the Optical Society of America, 4*, 1155-1167.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-145.

Edelman, S., & Weinshall, D. (1991). A self-organizing multipleview representation of three-dimensional objects. *Biological Cybernetics, 64*, 209-219.

Huttenlocher, D. P. (1988). *Three-dimensional recognition of solid objects from a two dimensional image.* Ph. D. Dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Johnson, S. H. (1993). *Recognizing Novel Views of Three-Dimensional Objects.* Cornell University manuscript.

Koenderink, J.J., & van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision.*Biological Cybernetics, 32*, 211-216.

Lowe, D.G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence, 31*, 355-395.

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

Perret, D. L., & Harries, M, H. (1988). Characteristic views and the visual inspection of simple faceted and smooth objects: 'tetrahedra and potatoes. *Perception,17*, 703-720.

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

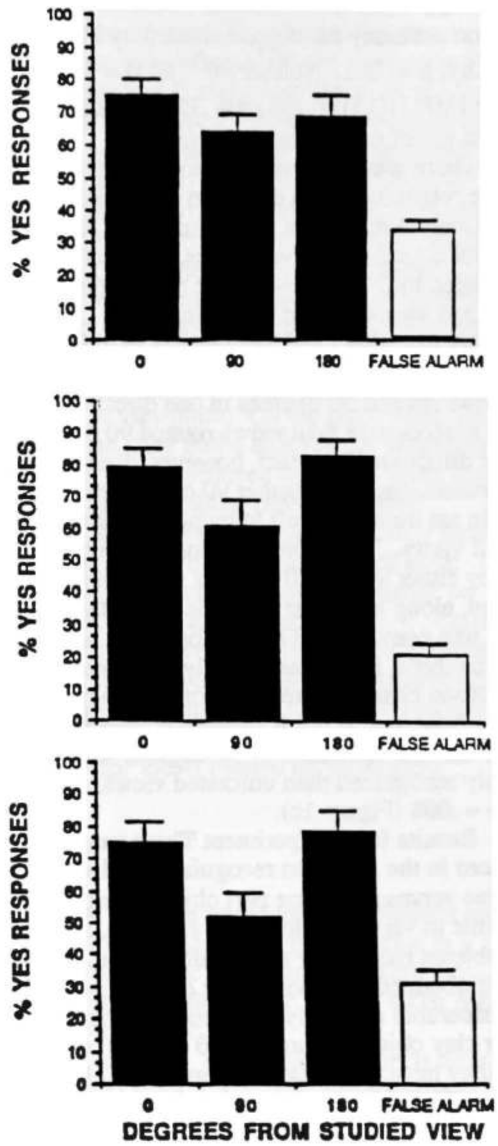
Rock, I., Di Vita, J., & Barbeito, R. (1981). The effect onform perception of change of orientation in the third dimension. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 719-732.

Seibert, M, & Waxman, A. M. (1992). Adaptive 3D-Object Recognition from Multiple Views. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Forthcoming.

Tarr, M.J. (1989). *Orientation-dependence in three-dimensional object recognition.* Ph. D. Dissertation, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, .

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

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32*, 193-254.



Figures 1a, 1b and 1c. Percentage of correct yes responses (i.e., hits) in Experiments One, Two and Three for familiar views (0^0), and novel views rotated by either 90^0 or 180^0.